



Calhoun: The NPS Institutional Archive

Theses and Dissertations

Thesis Collection

2007-03

Automatic author profiling of online chat logs

Lin, Jane.

Monterey, California. Naval Postgraduate School

<http://hdl.handle.net/10945/3559>



Calhoun is a project of the Dudley Knox Library at NPS, furthering the precepts and goals of open government and government transparency. All information contained herein has been approved for release by the NPS Public Affairs Officer.

Dudley Knox Library / Naval Postgraduate School
411 Dyer Road / 1 University Circle
Monterey, California USA 93943

<http://www.nps.edu/library>



**NAVAL
POSTGRADUATE
SCHOOL**

MONTEREY, CALIFORNIA

THESIS

**AUTOMATIC AUTHOR PROFILING OF ONLINE CHAT
LOGS**

by

Jane Lin

March 2007

Thesis Advisor:
Second Reader:

Craig H. Martell
Kevin M. Squire

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE March 2007	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE Automatic Author Profiling of Online Chat Logs			5. FUNDING NUMBERS	
6. AUTHOR(S) Jane Lin				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (maximum 200 words) <p>Now that the Internet has become easily accessible and more affordable, a larger number of people spend more time in front of a computer. Some spend so much time on the Internet that they develop virtual friendships and relationships - people with whom they have regular contact via a computer screen and the Internet. While most of the dialogue exchanged online is not harmful or illegal, there are those with dishonest intentions lurking online. These people can be breaking the law by seducing a minor virtually or even going as far as meeting a minor in person. Terrorists can also use the Internet to facilitate communication and plan attacks. Since e-mail is one of the original means of communication on the Internet, methods for determining the author of an email have already been studied. So far, however, no significant experimentation with online chat logs exists.</p> <p>The first of part of this study is comprised of generating an unbiased, random, and broad corpus of online chat logs. Having a general corpus with a wide-range of topics allows the results of this research to be applied in the most general case. Because developing a complete solution to the authorship attribution problem for chat logs is difficult, we limit our scope to predicting gender and age. The ultimate goal of this work, then, is to facilitate the jobs of law enforcers in tracking down criminals who attempt to use the Internet as a hiding place.</p>				
14. SUBJECT TERMS Authorship Analysis, Authorship Profiling, Authorship Attribution, Natural Language Processing, Naïve Bayes, Artificial Intelligence, Forensics, Information Retrieval, Data Mining			15. NUMBER OF PAGES 289	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

AUTOMATIC AUTHOR PROFILING OF ONLINE CHAT LOGS

Jane Lin
Civilian, Department of Defense
B.S. Cum Laude, University of Maryland, College Park, 2004

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

**NAVAL POSTGRADUATE SCHOOL
March 2007**

Author: Jane Lin

Approved by: Craig H. Martell, PhD
Thesis Advisor

Kevin M. Squire, PhD
Second Reader

Peter J. Denning, PhD
Chairman, Department of computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Now that the Internet has become easily accessible and more affordable, a larger number of people spend more time in front of a computer. Some spend so much time on the Internet that they develop virtual friendships and relationships - people with whom they have regular contact via a computer screen and the Internet. While most of the dialogue exchanged online is not harmful or illegal, there are those with dishonest intentions lurking online. These people can be breaking the law by seducing a minor virtually or even going as far as meeting a minor in person. Terrorists can also use the Internet to facilitate communication and plan attacks. Since e-mail is one of the original means of communication on the Internet, methods for determining the author of an email have already been studied. So far, however, no significant experimentation with online chat logs exists.

The first of part of this study is comprised of generating an unbiased, random, and broad corpus of online chat logs. Having a general corpus with a wide-range of topics allows the results of this research to be applied in the most general case. Because developing a complete solution to the authorship attribution problem for chat logs is difficult, we limit our scope to predicting gender and age. The ultimate goal of this work, then, is to facilitate the jobs of law enforcers in tracking down criminals who attempt to use the Internet as a hiding place.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION	1
A.	THE EARLY YEARS	1
B.	CHAT TODAY	1
C.	ROLE OF CHAT	2
D.	IMPORTANCE OF THE STUDY OF CHAT BEHAVIOUR	4
1.	Motivation	4
2.	Purpose	7
E.	ORGANIZATION OF THESIS	8
F.	CHAPTER SUMMARY	9
II.	BACKGROUND	11
A.	AUTHORSHIP ATTRIBUTION	11
B.	SOCIOLOINGUISTICS	11
C.	MACHINE LEARNING TECHNIQUES	12
1.	Naïve Bayes	13
2.	Support Vector Machines (SVM)	13
D.	CHAPTER SUMMARY	13
III.	GENERATION OF CORPUS	15
A.	SOURCE OF DATA	15
1.	Overview	15
2.	Chosen Host	15
B.	DATA DESCRIPTION	16
C.	STATISTICAL ANALYSIS OF DATA	17
1.	Document Counts, Sentence and Document Lengths	17
2.	Vocabulary Variety	25
3.	Emoticons	25
4.	Punctuations	28
D.	CHAPTER SUMMARY	30
IV.	TESTING AND ANALYSIS	31
A.	MACHINE LEARNING AND CLASSIFICATION TECHNIQUES	31
1.	Classification Tool	31
2.	Classification Method	32
3.	Measures of Classification Performance	32
B.	FEATURES AND FEATURE VECTORS	36
C.	EXPERIMENT SETUP	37
1.	Creating the Joint Probability Distribution Tables from the Training Data	37
2.	Labeling the Test Data	39
a.	Alternate Step 2: Finding the Best Label for Individual Features with Naïve Bayes	41

b.	<i>More Alternate Step 2's: Finding the Best Label for Feature Vectors with Naïve Bayes without the influence of the Prior</i>	42
D.	RESULTS AND ANALYSIS	42
1.	Including the Influence of the Prior	43
2.	Excluding the Influence of the Prior	46
E.	CHAPTER SUMMARY	50
V.	CONCLUSIONS	53
A.	SUMMARY	53
B.	FUTURE WORK	54
C.	CHAPTER SUMMARY AND CONCLUDING REMARKS	56
APPENDIX A:	TABLES FOR RAW COUNTS, FILE SIZES, SENTENCE LENGTHS, AND DOCUMENT LENGTHS	57
A.	RAW COUNTS FOR THE TRAINING SET	57
B.	RAW COUNTS FOR THE TESTING SET	60
APPENDIX B:	FIGURES OF SENTENCE LENGTHS AND DOCUMENT LENGTHS	61
A.	MEASURES OF AGE GROUP VS. SENTENCE LENGTH	61
B.	MEASURES OF INDIVIDUAL DOCUMENT LENGTHS	69
APPENDIX C:	TABLES FOR VOCABULARY VARIETY	71
A.	TYPE COUNTS	71
B.	AVERAGE VOCABULARY VARIETY	72
APPENDIX D:	TABLES AND FIGURES FOR EMOTICONS	73
A.	TABLES	73
B.	FIGURES	76
APPENDIX E:	TABLES AND FIGURES FOR PUNCTUATIONS	79
A.	TABLES	79
B.	FIGURES	86
APPENDIX F:	PRECISION, RECALL, AND F-SCORES FOR THE FEATURE VECTORS	91
A.	GENDER: BINARY CLASSIFICATION WITH PRIOR	91
1.	All Test Data	91
2.	Extracted Test Data: Teens and 20s	92
3.	Extracted Test Data: Teens and 30s	92
4.	Extracted Test Data: Teens and 40s	93
5.	Extracted Test Data: Teens and 50s	93
6.	Extracted Test Data: 20s and 30s	94
7.	Extracted Test Data: 20s and 40s	94
8.	Extracted Test Data: 20s and 50s	95
9.	Extracted Test Data: 30s and 40s	95
10.	Extracted Test Data: 30s and 50s	96
11.	Extracted Test Data: 40s and 50s	96

	12.	Extracted Test Data: Under 26 and 26 or Over	.97
B.		AGE: MULTI-CLASS (5-WAY) CLASSIFICATION WITH PRIOR	.97
	1.	All Test Data	.97
C.		AGE: BINARY CLASSIFICATION WITH PRIOR	.98
	1.	Extracted Test Data: Teens and 20s	.98
	2.	Extracted Test Data: Teens and 30s	.98
	3.	Extracted Test Data: Teens and 40s	.99
	4.	Extracted Test Data: Teens and 50s	.99
	5.	Extracted Test Data: 20s and 30s	100
	6.	Extracted Test Data: 20s and 40s	100
	7.	Extracted Test Data: 20s and 50s	101
	8.	Extracted Test Data: 30s and 40s	101
	9.	Extracted Test Data: 30s and 50s	101
	10.	Extracted Test Data: 40s and 50s	102
	11.	Extracted Test Data: Under 26 and 26 or Over	102
D.		GENDER: BINARY CLASSIFICATION WITHOUT PRIOR	103
	1.	All Test Data	103
	2.	Extracted Test Data: Teens and 20s	104
	3.	Extracted Test Data: Teens and 30s	104
	4.	Extracted Test Data: Teens and 40s	105
	5.	Extracted Test Data: Teens and 50s	105
	6.	Extracted Test Data: 20s and 30s	106
	7.	Extracted Test Data: 20s and 40s	106
	8.	Extracted Test Data: 20s and 50s	107
	9.	Extracted Test Data: 30s and 40s	107
	10.	Extracted Test Data: 30s and 50s	108
	11.	Extracted Test Data: 40s and 50s	108
	12.	Extracted Test Data: Under 26 and 26 or Over	109
E.		AGE: MULTI-CLASS (5-WAY) CLASSIFICATION WITHOUT PRIOR	110
	1.	All Test Data	110
F.		AGE: BINARY CLASSIFICATION WITHOUT PRIOR	111
	1.	Extracted Test Data: Teens and 20s	111
	2.	Extracted Test Data: Teens and 30s	111
	3.	Extracted Test Data: Teens and 40s	112
	4.	Extracted Test Data: Teens and 50s	112
	5.	Extracted Test Data: 20s and 30s	113
	6.	Extracted Test Data: 20s and 40s	113
	7.	Extracted Test Data: 20s and 50s	114
	8.	Extracted Test Data: 30s and 40s	114
	9.	Extracted Test Data: 30s and 50s	115
	10.	Extracted Test Data: 40s and 50s	115
	11.	Extracted Test Data: Under 26 and 26 or Over	116
APPENDIX G: PRECISION, RECALL, AND F-SCORES FOR THE			
INDIVIDUAL FEATURES			
A.		GENDER: BINARY CLASSIFICATION WITH PRIOR	117

	1.	All Test Data	117
	2.	Extracted Test Data: Teens and 20s	120
	3.	Extracted Test Data: Teens and 30s	123
	4.	Extracted Test Data: Teens and 40s	126
	5.	Extracted Test Data: Teens and 50s	129
	6.	Extracted Test Data: 20s and 30s	132
	7.	Extracted Test Data: 20s and 40s	135
	8.	Extracted Test Data: 20s and 50s	138
	9.	Extracted Test Data: 30s and 40s	141
	10.	Extracted Test Data: 30s and 50s	144
	11.	Extracted Test Data: 40s and 50s	147
	12.	Extracted Test Data: Under 26 and 26 or Over	150
B.		AGE: MULTI-CLASS (5-WAY) CLASSFICATION WITH PRIOR	152
	1.	All Test Data	152
C.		AGE: BINARY CLASSIFICATION WITH PRIOR	155
	1.	Extracted Test Data: Teens and 20s	155
	2.	Extracted Test Data: Teens and 30s	158
	3.	Extracted Test Data: Teens and 40s	161
	4.	Extracted Test Data: Teens and 50s	164
	5.	Extracted Test Data: 20s and 30s	166
	6.	Extracted Test Data: 20s and 40s	169
	7.	Extracted Test Data: 20s and 50s	172
	8.	Extracted Test Data: 30s and 40s	174
	9.	Extracted Test Data: 30s and 50s	177
	10.	Extracted Test Data: 40s and 50s	179
	11.	Extracted Test Data: Under 26 and 26 or Over	182
D.		GENDER: BINARY CLASSIFICATION WITHOUT PRIOR	185
	1.	All Test Data	185
	2.	Extracted Test Data: Teens and 20s	188
	3.	Extracted Test Data: Teens and 30s	191
	4.	Extracted Test Data: Teens and 40s	194
	5.	Extracted Test Data: Teens and 50s	197
	6.	Extracted Test Data: 20s and 30s	200
	7.	Extracted Test Data: 20s and 40s	203
	8.	Extracted Test Data: 20s and 50s	206
	9.	Extracted Test Data: 30s and 40s	209
	10.	Extracted Test Data: 30s and 50s	212
	11.	Extracted Test Data: 40s and 50s	215
	12.	Extracted Test Data: Under 26 and 26 or Over	218
E.		AGE: MULTI-CLASS (5-WAY) CLASSFICATION WITHOUT PRIOR	221
	1.	All Test Data	221
F.		AGE: BINARY CLASSIFICATION WITHOUT PRIOR	225
	1.	Extracted Test Data: Teens and 20s	225
	2.	Extracted Test Data: Teens and 30s	228
	3.	Extracted Test Data: Teens and 40s	231

4.	Extracted Test Data: Teens and 50s	234
5.	Extracted Test Data: 20s and 30s	237
6.	Extracted Test Data: 20s and 40s	240
7.	Extracted Test Data: 20s and 50s	243
8.	Extracted Test Data: 30s and 40s	246
9.	Extracted Test Data: 30s and 50s	249
10.	Extracted Test Data: 40s and 50s	252
11.	Extracted Test Data: Under 26 and 26 or Over 255	
APPENDIX H: KEY FOR FEATURE VECTORS AND FEATURES		259
A.	KEY FOR FEATURE VECTORS	259
B.	KEY FOR FEATURES	259
LIST OF REFERENCES		261
INITIAL DISTRIBUTION LIST		265

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 3-1.	Example of three sentences.....	20
Figure 3-2.	Age Group vs. Average Sentence Length.....	20
Figure 3-3.	Age Group vs. Average Document Length.....	21
Figure 3-4.	Age vs. Sentence Length for Males.....	22
Figure 3-5.	Age vs. Sentence Length for Females.....	23
Figure 3-6.	Age vs. Document Length in Tokens for Males....	24
Figure 3-7.	Age vs. Document Length in Tokens for Females..	24
Figure 3-8.	Age Group vs. Vocabulary Variety.....	25
Figure 3-9.	Age Group vs. Average Emoticon Token per Male Sentence.....	27
Figure 3-10.	Age Group vs. Average Emoticon Token per Female Sentence.....	27
Figure 3-11.	Age Group vs. Average Punctuation Tokens per Male Sentence.....	29
Figure 3-12.	Age Group vs. Average Punctuation Tokens per Female Sentence.....	29
Figure 4-1.	Two-Way Confusion Matrix.....	33
Figure 4-2.	Five-Way Confusion Matrix.....	34
Figure 4-3.	Step 1 - Counting Tokens and Types.....	38
Figure 4-4.	Step 2 - Binning into n bins.....	38
Figure 4-5.	Step 3 - Smoothing with Witten-Bell Discounting [17].....	39
Figure 4-6.	Step 1 - Calculating the Bins.....	40
Figure 4-7.	Step 2 - Finding the Best Label Using Naïve Bayes.....	41
Figure B-1.	Age Bin (Order 1) vs. Average Sentence Length..	61
Figure B-2.	Age Bin (Order 2) vs. Average Sentence Length..	62
Figure B-3.	Age Bin (Order 3) vs. Average Sentence Length..	63
Figure B-4.	Age Bin (Order 4) vs. Average Sentence Length..	64
Figure B-5.	Age Bin (Order 5) vs. Average Sentence Length..	65
Figure B-6.	Age Bin (Order 6) vs. Average Sentence Length..	66
Figure B-7.	Age Bin (Order 7) vs. Average Sentence Length..	67
Figure B-8.	Age Bin (Order 8) vs. Average Sentence Length..	68
Figure B-9.	Age vs. Document Length in Tokens for Males....	69
Figure B-10.	Age vs. Document Length in Tokens for Females..	69
Figure D-1.	Age Group vs. Average Emoticon Token per Sentence.....	77
Figure D-2.	Age Group vs. Emoticon Types per Sentence.....	77
Figure D-3.	Age Group vs. Average Emoticon Type per Male Sentence.....	78
Figure D-4.	Age Group vs. Average Emoticon Type per Female Sentence.....	78

Figure E-1.	Age Group vs. Average Punctuation Token per Sentence.....	87
Figure E-2.	Age Group vs. Average Punctuation Types per Sentence.....	88
Figure E-3.	Age Group vs. Average Punctuation Type per Male Sentence.....	88
Figure E-4.	Age Group vs. Average Punctuation Type per Female Sentence.....	89

LIST OF TABLES

Table 1-1.	Characteristics of Offenders who were Arrested for Internet Sex Crimes Against Minors [From Ref. 16].....	7
Table 3-1.	Count of Screen Names for the Training Set.....	18
Table 3-2.	Count of Screen Names for the Test Set.....	18
Table 3-3.	Average Token Count for the Training Set.....	19
Table 3-4.	Built-In Emoticons.....	26
Table 4-1.	Extracted Features and Feature Vectors for Males Including the Prior.....	44
Table 4-2.	Extracted Features and Feature Vectors for Females Including the Prior.....	44
Table 4-3.	Extracted Features and Feature Vectors for Teens Including the Prior.....	45
Table 4-4.	Extracted Features and Feature Vectors for 20s Including the Prior.....	45
Table 4-5.	Extracted Features and Feature Vectors for 30s Including the Prior.....	45
Table 4-6.	Extracted Features and Feature Vectors for 40s Including the Prior.....	45
Table 4-7.	Extracted Features and Feature Vectors for 50s Including the Prior.....	46
Table 4-8.	Extracted Features and Feature Vectors for Under 26 Including the Prior.....	46
Table 4-9.	Extracted Features and Feature Vectors for 26 and Over Including the Prior.....	46
Table 4-10.	Extracted Features and Feature Vectors for Males Excluding the Prior.....	47
Table 4-11.	Extracted Features and Feature Vectors for Females Excluding the Prior.....	48
Table 4-12.	Extracted Features and Feature Vectors for Teens Excluding the Prior.....	48
Table 4-13.	Extracted Features and Feature Vectors for 20s Excluding the Prior.....	49
Table 4-14.	Extracted Features and Feature Vectors for 30s Excluding the Prior.....	49
Table 4-15.	Extracted Features and Feature Vectors for 40s Excluding the Prior.....	49
Table 4-16.	Extracted Features and Feature Vectors for 50s Excluding the Prior.....	50
Table 4-17.	Extracted Features and Feature Vectors for Under 26 Excluding the Prior.....	50
Table 4-18.	Extracted Features and Feature Vectors for 26 and Over Excluding the Prior.....	50

Table A-1.	Token Count for Training Set.....	57
Table A-2.	Sentence Count for Training Set.....	58
Table A-3.	File Size for Training Set.....	58
Table A-4.	Average Sentence Length (Tokens/Sentence).....	59
Table A-5.	Average Document Length (Tokens/Documents).....	59
Table A-6.	File Size for Testing Set.....	60
Table C-1.	Type Count for Training Set.....	71
Table C-2.	File Size for Testing Set.....	72
Table D-1.	Emoticon Token per Male Sentence.....	73
Table D-2.	Emoticon Token per Female Sentence.....	74
Table D-3.	Emoticon Type per Male Sentence.....	75
Table D-4.	Emoticon Type per Female Sentence.....	76
Table E-1.	Punctuation Tokens per Male Sentence.....	80
Table E-2.	Average Punctuation Tokens per Female Sentence.....	82
Table E-3.	Average Punctuation Types per Male Sentence.....	84
Table E-4.	Average Punctuation Types per Female Sentence.....	86
Table F-1.	P, R, F-Score for Males.....	91
Table F-2.	P, R, F-Score for Females.....	91
Table F-3.	P, R, F-Score for Males.....	92
Table F-4.	P, R, F-Score for Females.....	92
Table F-5.	P, R, F-Score for Males.....	92
Table F-6.	P, R, F-Score for Females.....	92
Table F-7.	P, R, F-Score for Males.....	93
Table F-8.	P, R, F-Score for Females.....	93
Table F-9.	P, R, F-Score for Males.....	93
Table F-10.	P, R, F-Score for Females.....	93
Table F-11.	P, R, F-Score for Males.....	94
Table F-12.	P, R, F-Score for Females.....	94
Table F-13.	P, R, F-Score for Males.....	94
Table F-14.	P, R, F-Score for Females.....	94
Table F-15.	P, R, F-Score for Males.....	95
Table F-16.	P, R, F-Score for Females.....	95
Table F-17.	P, R, F-Score for Males.....	95
Table F-18.	P, R, F-Score for Females.....	95
Table F-19.	P, R, F-Score for Males.....	96
Table F-20.	P, R, F-Score for Females.....	96
Table F-21.	P, R, F-Score for Males.....	96
Table F-22.	P, R, F-Score for Females.....	96
Table F-23.	P, R, F-Score for Males.....	97
Table F-24.	P, R, F-Score for Females.....	97
Table F-25.	P, R, F-Score for Teens.....	97
Table F-26.	P, R, F-Score for 20s.....	97
Table F-27.	P, R, F-Score for Teens.....	98
Table F-28.	P, R, F-Score for 20s.....	98
Table F-29.	P, R, F-Score for Teens.....	98
Table F-30.	P, R, F-Score for 20s.....	98

Table F-31.	P, R, F-Score for Teens.....	99
Table F-32.	P, R, F-Score for 40s.....	99
Table F-33.	P, R, F-Score for Teens.....	99
Table F-34.	P, R, F-Score for 20s.....	100
Table F-35.	P, R, F-Score for 30s.....	100
Table F-36.	P, R, F-Score for 20s.....	100
Table F-37.	P, R, F-Score for 40s.....	100
Table F-38.	P, R, F-Score for 20s.....	101
Table F-39.	P, R, F-Score for 30s.....	101
Table F-40.	P, R, F-Score for 40s.....	101
Table F-41.	P, R, F-Score for 30s.....	101
Table F-42.	P, R, F-Score for 40s.....	102
Table F-43.	P, R, F-Score for 50s.....	102
Table F-44.	P, R, F-Score for Under 26.....	102
Table F-45.	P, R, F-Score for 26 or older.....	102
Table F-46.	P, R, F-Score for Males.....	103
Table F-47.	P, R, F-Score for Females.....	103
Table F-48.	P, R, F-Score for Males.....	104
Table F-49.	P, R, F-Score for Females.....	104
Table F-50.	P, R, F-Score for Males.....	104
Table F-51.	P, R, F-Score for Females.....	104
Table F-52.	P, R, F-Score for Males.....	105
Table F-53.	P, R, F-Score for Females.....	105
Table F-54.	P, R, F-Score for Males.....	105
Table F-55.	P, R, F-Score for Females.....	105
Table F-56.	P, R, F-Score for Males.....	106
Table F-57.	P, R, F-Score for Females.....	106
Table F-58.	P, R, F-Score for Males.....	106
Table F-59.	P, R, F-Score for Females.....	106
Table F-60.	P, R, F-Score for Males.....	107
Table F-61.	P, R, F-Score for Females.....	107
Table F-62.	P, R, F-Score for Males.....	107
Table F-63.	P, R, F-Score for Females.....	107
Table F-64.	P, R, F-Score for Males.....	108
Table F-65.	P, R, F-Score for Females.....	108
Table F-66.	P, R, F-Score for Males.....	108
Table F-67.	P, R, F-Score for Females.....	108
Table F-68.	P, R, F-Score for Males.....	109
Table F-69.	P, R, F-Score for Females.....	109
Table F-70.	P, R, F-Score for Teens.....	110
Table F-71.	P, R, F-Score for 20s.....	110
Table F-72.	P, R, F-Score for 30s.....	110
Table F-73.	P, R, F-Score for 40s.....	110
Table F-74.	P, R, F-Score for 50s.....	110
Table F-75.	P, R, F-Score for Teens.....	111
Table F-76.	P, R, F-Score for 20s.....	111

Table F-77.	P, R, F-Score for Teens.....	111
Table F-78.	P, R, F-Score for 30s.....	111
Table F-79.	P, R, F-Score for Teens.....	112
Table F-80.	P, R, F-Score for 40s.....	112
Table F-81.	P, R, F-Score for Teens.....	112
Table F-82.	P, R, F-Score for 50s.....	112
Table F-83.	P, R, F-Score for 20s.....	113
Table F-84.	P, R, F-Score for 30s.....	113
Table F-85.	P, R, F-Score for 20s.....	113
Table F-86.	P, R, F-Score for 40s.....	113
Table F-87.	P, R, F-Score for 20s.....	114
Table F-88.	P, R, F-Score for 50s.....	114
Table F-89.	P, R, F-Score for 30s.....	114
Table F-90.	P, R, F-Score for 40s.....	114
Table F-91.	P, R, F-Score for 30s.....	115
Table F-92.	P, R, F-Score for 50s.....	115
Table F-93.	P, R, F-Score for 30s.....	115
Table F-94.	P, R, F-Score for 40s.....	115
Table F-95.	P, R, F-Score for under 26.....	116
Table F-96.	P, R, F-Score for 26 and Older.....	116
Table G-1.	P, R, F-Score for Males.....	119
Table G-2.	P, R, F-Score for Females.....	119
Table G-3.	P, R, F-Score for Males.....	120
Table G-4.	P, R, F-Score for Females.....	122
Table G-5.	P, R, F-Score for Males.....	123
Table G-6.	P, R, F-Score for Females.....	125
Table G-7.	P, R, F-Score for Males.....	126
Table G-8.	P, R, F-Score for Females.....	128
Table G-9.	P, R, F-Score for Males.....	129
Table G-10.	P, R, F-Score for Females.....	131
Table G-11.	P, R, F-Score for Males.....	133
Table G-12.	P, R, F-Score for Females.....	134
Table G-13.	P, R, F-Score for Males.....	136
Table G-14.	P, R, F-Score for Females.....	137
Table G-15.	P, R, F-Score for Males.....	139
Table G-16.	P, R, F-Score for Females.....	140
Table G-17.	P, R, F-Score for Males.....	142
Table G-18.	P, R, F-Score for Females.....	143
Table G-19.	P, R, F-Score for Males.....	145
Table G-20.	P, R, F-Score for Females.....	146
Table G-21.	P, R, F-Score for Males.....	148
Table G-22.	P, R, F-Score for Females.....	149
Table G-23.	P, R, F-Score for Males.....	151
Table G-24.	P, R, F-Score for Females.....	152
Table G-25.	P, R, F-Score for Teens.....	152
Table G-26.	P, R, F-Score for Males.....	154

Table G-27.	P, R, F-Score for Teens.....	155
Table G-28.	P, R, F-Score for 20s.....	157
Table G-29.	P, R, F-Score for Teens.....	159
Table G-30.	P, R, F-Score for 30s.....	160
Table G-31.	P, R, F-Score for Teens.....	162
Table G-32.	P, R, F-Score for 40s.....	163
Table G-33.	P, R, F-Score for Teens.....	165
Table G-34.	P, R, F-Score for 20s.....	167
Table G-35.	P, R, F-Score for 30s.....	168
Table G-36.	P, R, F-Score for 20s.....	170
Table G-37.	P, R, F-Score for 40s.....	171
Table G-38.	P, R, F-Score for 20s.....	173
Table G-39.	P, R, F-Score for 30s.....	175
Table G-40.	P, R, F-Score for 40s.....	176
Table G-41.	P, R, F-Score for 30s.....	178
Table G-42.	P, R, F-Score for 40s.....	180
Table G-43.	P, R, F-Score for 40s.....	181
Table G-44.	P, R, F-Score for Under 26.....	183
Table G-45.	P, R, F-Score for 26 and Older.....	184
Table G-46.	P, R, F-Score for Males.....	186
Table G-47.	P, R, F-Score for Females.....	187
Table G-48.	P, R, F-Score for Males.....	188
Table G-49.	P, R, F-Score for Females.....	190
Table G-50.	P, R, F-Score for Males.....	191
Table G-51.	P, R, F-Score for Females.....	193
Table G-52.	P, R, F-Score for Males.....	194
Table G-53.	P, R, F-Score for Females.....	196
Table G-54.	P, R, F-Score for Males.....	197
Table G-55.	P, R, F-Score for Females.....	199
Table G-56.	P, R, F-Score for Males.....	201
Table G-57.	P, R, F-Score for Females.....	202
Table G-58.	P, R, F-Score for Males.....	204
Table G-59.	P, R, F-Score for Females.....	205
Table G-60.	P, R, F-Score for Males.....	207
Table G-61.	P, R, F-Score for Females.....	208
Table G-62.	P, R, F-Score for Males.....	210
Table G-63.	P, R, F-Score for Females.....	211
Table G-64.	P, R, F-Score for Males.....	213
Table G-65.	P, R, F-Score for Females.....	214
Table G-66.	P, R, F-Score for Males.....	216
Table G-67.	P, R, F-Score for Females.....	217
Table G-68.	P, R, F-Score for Males.....	219
Table G-69.	P, R, F-Score for Females.....	221
Table G-70.	P, R, F-Score for Teens.....	221
Table G-71.	P, R, F-Score for 20s.....	223
Table G-72.	P, R, F-Score for 30s.....	223

Table G-73.	P, R, F-Score for 40s.....	224
Table G-74.	P, R, F-Score for 50s.....	224
Table G-75.	P, R, F-Score for Teens.....	225
Table G-76.	P, R, F-Score for 20s.....	227
Table G-77.	P, R, F-Score for Teens.....	229
Table G-78.	P, R, F-Score for 30s.....	230
Table G-79.	P, R, F-Score for Teens.....	232
Table G-80.	P, R, F-Score for 40s.....	233
Table G-81.	P, R, F-Score for Teens.....	235
Table G-82.	P, R, F-Score for 50s.....	236
Table G-83.	P, R, F-Score for 20s.....	238
Table G-84.	P, R, F-Score for 30s.....	239
Table G-85.	P, R, F-Score for 20s.....	241
Table G-86.	P, R, F-Score for 40s.....	242
Table G-87.	P, R, F-Score for 20s.....	244
Table G-88.	P, R, F-Score for 50s.....	245
Table G-89.	P, R, F-Score for 30s.....	247
Table G-90.	P, R, F-Score for 40s.....	248
Table G-91.	P, R, F-Score for 30s.....	250
Table G-92.	P, R, F-Score for 50s.....	251
Table G-93.	P, R, F-Score for 40s.....	253
Table G-94.	P, R, F-Score for 50s.....	254
Table G-95.	P, R, F-Score for Under 26.....	256
Table G-96.	P, R, F-Score for 26 or Older.....	257
Table H-1.	Key for Feature Vectors in Appendix F.....	259
Table H-2.	Key for Features in Appendix G.....	260

ACKNOWLEDGMENTS

I owe my success to the help of many people. Without their help and encouragement, I would not be where I am today. Therefore, I would like to acknowledge the following people:

Craig Martell, thank you for pushing me into finding something that interests me, providing your guidance and expertise, and keeping me on track (while encouraging me go snowboard often) as my thesis advisor.

Kevin Squire, thank you for your guidance, interest in my work, and technical discussions when I needed someone to look at the problem from a different angle.

Eric Forsyth, thank you for helping me scrub my data. Without your help, I may not have been able to finish on time.

Paul Avellino, thank you for your support and encouragement throughout the quarters, especially near graduation time.

Prof. Luqi, thank you for your curiosity in my work and staying in touch throughout my career at the Naval Postgraduate School to make sure that I was progressing.

Prof. Irvine, thank you for your help with my career both at the Naval Postgraduate School and after. Your stories in class were one of a kind and your professional contacts are invaluable.

Prof. Reihle, thank you for your advice and encouragement during my time at the Naval Postgraduate

School. Your spirit and excitement for Judo is very inspiring and motivating.

My parents, thank you for your constant encouragement to pursue my goals and dreams in life. Without your guidance and support, I may not have come to the Naval Postgraduate School in the first place where I had an unforgettable, eye-opening, and unique learning experience.

In addition to the above, I would also like to extend my acknowledgements to all my other colleagues at the Naval Postgraduate School, especially those in the Scholarship for Service program and the Graduate Training Program, thank you all for providing me an outlet away from all the work and stress.

I. INTRODUCTION

A. THE EARLY YEARS

Like everything else, chat has evolved over the years. Early time-sharing systems in the 1960's were designed to support real-time chat. However, not only did users need to be connected to the same system, the chat program could only transmit messages between two users at one time. One of the best known early chat utilities is the "talk" function on Unix operating systems. Initially, talk could only facilitate communication amongst users on a single multi-user computer. Later on, talk expanded to allow communication across multiple computers as well. Variations and extensions of talk allowed for communication between more than two users in a restricted-broadcast manner where the sender had to specify the address of each recipient. With improvements in technology, chat continued to transform and be more realistic.

B. CHAT TODAY

The growth of the Internet fueled the rapid expansion of online chat. Developed in 1988 by Jarkko Oikarinen of Finland, the Internet Relay Chat (IRC) protocol led to the beginning of real-time chat between large groups of users located in different parts of the world. Through modifications and enhancements, modern IRC operates on a network of servers which relay messages to each other allowing users on one server to communicate with users on any of the other servers on the same network. These networks contain thousands of chat rooms (called "channels") with tens of thousands of users.

ICQ, AOL Messenger, and MSN Messenger are IRC-like chat programs that also include exchange of files via email or direct connection. Originally, these required software that must be downloaded and installed; only users with the necessary software could chat with each other. Because this makes it relatively easy for companies to block chat traffic, it is more common today to have web interfaces that allow connection to these and other chat servers.

While the features and methods of chat evolved over time, the fundamental purpose of chat still remain the same - allowing users to communicate real-time with one another without face-to-face contact.

C. ROLE OF CHAT

Because the internet is now an integral part of life, it is unsurprising that chat continues to grow as a means of communication among friends, family, colleagues, and even complete strangers. There was a time when family members who did not live in the same area needed to use a phone card or have a long distance provider in order to talk to their loved ones. This would often result in a very large phone bill for both parties. However, since using chat presupposes only having a network connection, many people have come to rely on chat as a means of staying in touch with long distance friends. Many chat hosts today (e.g. MSN, Yahoo!, AIM) allow video or voice conferencing to make it even more realistic and personal. This is all done without extra cost to either party, except for maybe a microphone or webcam.

The use of chat in business is pretty common today too. Some companies require their employees to install the company-wide messaging service and be logged on during work

hours. Because of how long it is to communicate through chat, many companies have set up security perimeters to disable chat outside of a company's network. This remedies the problem of employees spending too much time chatting with non-employees during work hours. On the other hand, having chat communication in the office saves employees time and can actually encourage more communication. For example, instead of walking from one end of the hall to the other and back to ask a simple question (such as "What time is meeting?"), it would take less than thirty-seconds to complete this dialogue over chat. Even more time would be wasted if the colleague was not in his/her office at the time. So, why not use a phone? Picking up the phone and calling could end up interrupting something. With chat, the colleague could respond at his/her earliest convenience without feeling obligated for immediate response. The message(s) will also stay on the colleague's computer, so the colleague does not have to constantly remind him/herself to remember to respond.

Chat is also an outlet for shy people or for people with self-image problems. Hidden behind a computer screen, the identity of the user is concealed allowing people to freely engage in conversation focusing on the topic of discussion rather than on appearances or other issues. As long as the internet is functioning, someone somewhere in the world is chatting. Chat never sleeps, thus making it a natural place to look for interaction with other people in the comfort of one's own home. Chat also facilitates communication between different types of people who normally would not converse with each other. While chat makes it possible to interact with a plethora of people, it

is also easy to ignore those who are rude or who do not provide good company.

D. IMPORTANCE OF THE STUDY OF CHAT BEHAVIOUR

1. Motivation

The growing number of sex crimes committed against children and youth has been fueled by the expansion of the Internet and the increasing commonality of computers [35]. In 1999, Dr. David Finkelhor, Director of the Crimes Against Children Research Center at the University of New Hampshire was funded by the National Center for Missing and Exploited Children (NCMEC) to conduct a research survey on Internet victimization of youth [16]. The project staff interviewed a nationally-representative sample of 1,501 youth between the ages of 10 and 17 who used the Internet regularly. "Regular use" was defined as using the Internet at least once a month for the past six months [16]. As defined by Finkelhor, the four types of online victimization of youth studied in this survey included:

- Sexual solicitation and approaches: Requests to engage in sexual activities or sexual talks or to give personal sexual information that were unwanted or, whether wanted or not, made by an adult.
- Aggressive sexual solicitation: Sexual solicitations involving offline contact with the perpetrator through mail, by telephone, or in person, or attempts or requests for offline contact. This also included a predator sending money or gifts through the U.S. postal Service to a young person.

- Unwanted exposure to sexual material: When online, opening email, or opening email links, and not seeking or expecting sexual material, being exposed to pictures of naked people or people having sex.
- Harassment: Threats or other offensive content (not sexual solicitation) sent online to the youth or posted online for others to see.

Interesting statistical highlights supporting the dangers of the Internet are found in Finkelhor's study. These include having:

- One in 5 youths received a sexual approach or solicitation over the Internet during that past year.
- One in 33 youths received an aggressive sexual solicitation in the past year.
- One in 4 youths received unwanted exposure in the past year to pictures of naked people or people having sex.
- One in 17 youths was threatened or harassed in the past year.

Finkelhor also surveyed responses of youths in these situations. He found that while about 25 percent of the youth who encountered a sexual approach or solicitation told a parent, almost 40 percent of those reporting an unwanted exposure to sexual material told a parent [16]. However, only 17 percent of youths and 11 percent of parents could name a specific authority, such as the Federal Bureau

of Investigation (FBI), CyberTipline, or an Internet service provider, to which they could report an Internet crime [16].

In 2003, Janis Wolak, Kimberly Mitchell, and David Finkelhor conducted the first research to gather statistics of offenders who were arrested for Internet sex crimes. Table 1-1 shows the summary of their findings.

OFFENDER CHARACTERISTICS	% (weighted n = 2,577)
Gender of Offender: Male	99%
Race of Offender: Non-Hispanic White	92%
Age of Offender:	
17 or Younger	3%
18 to 25	11%
26 to 39	45%
40 or Older	41%
Other Characteristics:	
Acted Alone in Crime	97%
Prior Arrests for Sexual Offending Against Minors	10%
Known to be Violent to any Degree	11%
Possessed Child Pornography	67%
Distributed Child Pornography	22%
Solicited an Undercover Investigator	27%
Committed a Sex Crime Against an Identified Victim	45%

Crime Against Identified Victim was:	
Internet-Initiated	20%
Against a Family Member or Prior Acquaintance of the Offender	19%
Not Internet-Related	7%

Table 1-1. Characteristics of Offenders who were Arrested for Internet Sex Crimes Against Minors [From Ref.

16]

Not only is the Internet becoming a popular place for child predators to hide, it can also facilitate communication for those with other kinds of unfavorable agendas. For example, online chat allows terrorists (just like it would any other persons) to have instant correspondence. This form of communication accelerates planning and enhances organization of terrorists and impedes the Global War on Terror.

2. Purpose

With the decreasing cost of access to the Internet combined with developing technologies, new challenges arise for law enforcement requiring them to confront situations not anticipated in criminal statutes, master technical advances, develop new investigative techniques, and handle criminal cases that often span multiple jurisdictions [35]. Since the vast majority of online sex offenders were non-Hispanic White males older than 25 who were acting alone (See Table 1-1), it would be extremely helpful to have an automatic way of identifying such parties. Similarly, once studies of profiles of terrorists who communicates via online chat is determined, many, including the governments of other nations, would be very interested in finding a way

for automatic detection or author attribution. However, developing a complete solution to this problem is very difficult (even if common characteristics are already discovered as in the case of sex offenders). Thus, the purpose of this thesis is to introduce author attribution of online chat logs since no significant experimentation with such data exists. Although we limit the scope of this thesis to predicting gender and age, the ultimate goal of this work is to facilitate the jobs of law enforces in tracking down criminals who attempt to use the Internet as a hiding place.

E. ORGANIZATION OF THESIS

This thesis is organized as follows:

- Chapter I provides an introduction of online chat history and role and the importance of studying chat followed by the motivation and purpose of the thesis.
- Chapter II provides a background of early and current authorship analysis and attribution and previous work done on sociolinguistics. Machine learning techniques such as Naïve Bayes and Support Vector Machines are also introduced.
- Chapter III provides details on the corpus generation since there is no chat corpus available. After describing the data collected, some statistical analysis of the data is also provided. Based on the statistical analysis, it appears that there may be some trends with specific features in distinguishing and/or gender. However, no conclusive remarks can be provided at this step.

- Chapter IV explains the machine learning tools and classification methods used in this research as well as the focus of features and feature vectors. After description of the experiment setup, the results are presented.
- Chapter V concludes this thesis with a summary of the goal and results, future work, and last remarks.
- The appendices follow with a listing of supporting tables and figures for sections through out the thesis.

F. CHAPTER SUMMARY

In this introductory chapter, we motivated this research by describing potential applications and benefits of automatic author profiling of online chats for law enforcers after giving a brief summary of the role of online chat. Next, we presented the organization of the thesis. We continue with the background of this research in Chapter II.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. AUTHORSHIP ATTRIBUTION

Authorship attribution dates back to the late nineteenth century with the studies of Mendenhall [22] and Mascol [20,21] to what is now called *stylometrics*. More than 40 years later, Yule [33,34] and Zipf [37] influenced the characteristic of the early work with their textual statistics, Yule's K statistic and Zipf's distribution, respectively. While work in this area initially focused on literature and gospels of the New Testament, modern work in authorship attribution, or non-traditional authorship attribution, began in 1964 with the study of *The Federalist Papers* by Mosteller and Wallace [23]. Authorship analysis of computer programs started in the late twentieth century by Gary, Sallis, and MacDonell [13] and Krsul and Spafford [18]. E-mail authorship began in the twenty-first century with the work of DeVel [7,8,9,10], his student Corney [5,6], and Argamon [1,2].

Unlike published literature, e-mail has a much smaller text sampling that can be used to generate precise language models. Chat logs further aggravate this problem because sampling is even shorter with the high possibility of multiple topics in alternating sentences.

B. SOCIOLOINGUISTICS

Although men and women speak the same language, empirical evidence suggests that women converse differently than men [5]. Ojemann found that different parts of the brain are activated by men and women for some language tasks [24]. Brizendine found that females talk three times more than males a day because a bigger portion of the

female brain is dedicated to speech [4]. Similarly, Singh found that male speech was lexically richer and tended to use longer phrases, while female speech used more verbs and shorter sentence structures [29].

Not only are there differences in gender, studies of differences in communication (i.e. speech, writing, web postings, discussion groups, email, web blogs) for different ages, social groups, educational levels, and language background have also been done [3,13,14,25,27,28,29,30,31]. Findings from [13] on web postings suggest that women use a *rapport* style of communication, but men use a *report* style of communication; women are more likely to express doubt, apologize, ask questions, suggest ideas rather where as men were more likely to show self promotion, make insults, use sarcasm, and make strong assertions. The experiment in [3] suggests that although there is no authorial structure differences between different educational level, there were some difference in measures of vocabulary richness. Being one of the newest forms of communication, very few studies involving chat have been done and no studies attempting to automate the classification of gender and age in general online chat logs have been found in the literature.

C. MACHINE LEARNING TECHNIQUES

Stylometry makes measures of the discriminatory features proposed for authorship attribution, which, in turn, reduces the style of a particular author's profile to a pattern [5]. A pattern matching problem is especially suited for machine learning. Machine learning allows the classification of unseen data by producing a model based on the knowledge it learned from previously seen data. The

performance of a machine learning algorithm must be measurable in order for the evaluator to determine the accuracy of the model. Various machine learning techniques have been tried in authorship attribution. Two of the more common ones include Naïve Bayes and Support Vector Machines [5,11,19,32].

1. Naïve Bayes

The Naïve Bayesian Classifier, or Naïve Bayes, is one of the most effective classification algorithms [36]. It is relatively simple to compute since it assumes independence between features. Naïve Bayes relies on the Bayesian model developed by the British mathematician, Thomas Bayes in 1763. Being a good first-step analysis tool, this technique is used in this research. Further discussion on Naïve Bayes is given in Section IV, A.

2. Support Vector Machines (SVM)

Support Vector Machines are becoming a more frequency used technique for authorship attribution. Developed by Vladimir Vapnik in 1995, SVM takes a set of features and performs some calculations to arrive in a new space where a hyperplane can be determined to split the feature vectors in the new space. The ideal hyperplane separates the feature vectors with maximum margins. Since there are only two sides to the hyperplane, this technique is best suited for binary problems. See [15] for use of SVMs in multi-class problems.

D. CHAPTER SUMMARY

In this chapter, the movement of author attribution is briefly described, and previous findings in sociolinguists are presented. An introduction of common machine learning techniques in authorship attribution (i.e. Naïve Bayes and Support Vector Machines) is also provided. In the next

chapter, we present how the corpus was generated along with some statistical analysis on the data.

III. GENERATION OF CORPUS

A. SOURCE OF DATA

1. Overview

The internet includes thousands of chat hosts. Some require users to pay a membership fee, while others offer free services with a simple registration. Still, there are others that do not require any sort of commitment or identifying information. Chat hosts range in purpose and intended audiences as well as in the variety of *rooms* they offer. Some are advertised as a way to meet singles, friends, other married persons, etc. Others advertise discussions in business, technology, health, relationships, religion, etc. Many popular email hosts such as Gmail, Yahoo!, or AOL also offer chat for its users. Note that this research does not sponsor and is not sponsored by any of the aforementioned chat hosts.

2. Chosen Host

A publicly available chat host has been used in this research to gather data. Important factors considered when choosing which chat host to use included its customer variety and chat content coverage. The chosen chat host runs on Java, so it is not platform specific. Thus, the data collected is from a more general group of people on different systems. While there are scheduled chat rooms available only at specific times, standard chat rooms available all the time and personal chat rooms created and customized by its users also exist. The standard chat rooms are comprised of rooms by a variety of topics. However, to keep the data as general and unbiased as

possible, the rooms chosen for this research included only standard rooms organized by age groups.

B. DATA DESCRIPTION

The length of the user's actual screen name is limited to twelve alphanumeric or underscore characters by the text box on the registration page. However, users can also create nicknames that do not have any of these restrictions. If a user changes their screen name to a nickname, others can only see the user by his/her nickname thereby allowing users to hide their "true" identity. This is not an issue for data collection because linking a nickname to the actual screen name is trivial when network packets are captured.

The message portion of the data includes users' messages, system messages, and bot messages. Bot messages were not included in this study because bots do not have age and gender characteristics. System messages welcoming a user to a room or informing a change in state of the room (i.e. a user entering or leaving the room) were also discarded because they are not part of the user's chat content or style. However, system messages signifying a change in font color, font size, or font style was preserved with the user's messages, since it is consciously triggered by the user and represents the user's style. Users' messages include words, abbreviations or acronyms, and emoticons.

Along with the messages, users' profile information (i.e. age and gender) was retrieved from the chat host database to complete the set of data. With the profile information associated with the chat messages, we now remove all appearances of the original screen names for

privacy protection of the users. Thus, no data retained in the corpus can be traced back to the original screen name or to any other identifying information besides the age and the gender. Only chat messages from users with gender information were retained for use in this research while all messages from other users were thrown out. Data from all users over the age of 60 was also discarded due to scarcity of data. From this smaller set of data, 90% was set aside as the training set while 10% was used as the testing set.

C. STATISTICAL ANALYSIS OF DATA

1. Document Counts, Sentence and Document Lengths

Because research on chat logs is a new area, gender and age distinguishing features are not yet understood. Part of this research is to find out what features are useful in determining whether the user is a male or female and to what age group he/she belongs. Thus, statistics are gathered from the training set to discover the screen name count, average sentence length, average token count, punctuation count, emoticon count, and vocabulary count for the different classifications.

Tables 3-1 and 3-2 show a summary of the screen name count separated by age groups and gender for both the training set and the test set, respectively. Each unique screen name constitutes one count. Thus, the two tables combined show that a total of 3289 unique screen names exist in the entire data set.

Age Group	Male	Female	Total
Unspecified	411	317	728
13-19	207	384	591
20-29	464	418	882
30-39	214	141	355
40-49	183	118	301
50-59	78	25	103
Total	1557	1403	2960

Table 3-1. Count of Screen Names for the Training Set

Age Group	Male	Female	Total
Unspecified	48	37	85
13-19	22	46	68
20-29	52	45	97
30-39	23	14	37
40-49	20	12	32
50-59	8	2	10
Total	173	156	329

Table 3-2. Count of Screen Names for the Test Set

Table 3-3 shows the average token count for the training data set. A token is defined as a contiguous string of characters that is surrounded by whitespace, the beginning of a sentence, or the end of a sentence. The

average token count is derived by counting up the number of tokens and dividing by the number of screen names shown in Table 2. For example, for males across all ages, 428522 tokens were found (see Appendix A). Dividing 428522 by 1557 yields approximately 275.22 as seen in the second column of the last row.

Age Group	Male	Female	Both
Unspecified	289.34	256.21	274.91
13-19	198.29	162.61	175.1
20-29	283.25	357.21	318.3
30-39	330.59	551.44	418.31
40-49	295.26	456.06	358.3
50-59	158.37	229.68	175.68
All Ages	275.22	306.69	290.14

Table 3-3. Average Token Count for the Training Set

Figure 3-1 shows an example of three sentences. A sentence is defined as a line of text preceded by the user's screen name. Note that even though the second sentence in Figure 3 wraps to the next line, it still counts as one sentence because the second line is not preceded by the user's screen name. This is used to generate Figure 3-2, which depicts the average sentence length. The average sentence length is defined as the average of the number of tokens per sentence and is calculated by dividing the number of tokens (Table A-1) by the number of sentences (Table A-2). Appendix A contains the table of actual values of the points on this graph.

Ages 13-19 were clumped together and represented as data point 15; ages 20-29 were clumped together and represented as data point 25; ages 30-39 were clumped together and represented as data point 35; ages 40-49 were clumped together and represented as data point 45; ages 50-59 were clumped together and represented as data point 55.

Bob: Hi Alice!

Alice: Hello, Bob. Long time no chat. How are your wife and kids doing?

Bob: They are great! Thanks for asking.

Figure 3-1. Example of three sentences

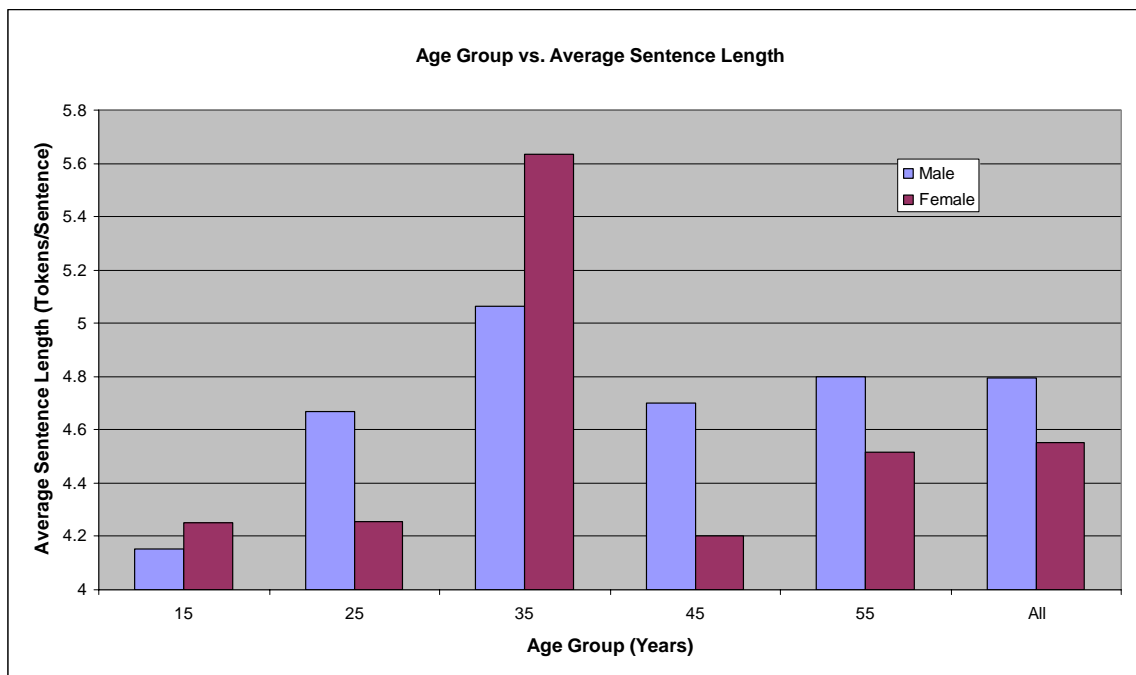


Figure 3-2. Age Group vs. Average Sentence Length

It is interesting to see the average sentence length peak in the 30's range for both male and females when the age groups are grouped this way. The last data point ("All") represents the average of all age groups for a particular gender. On average across all ages, it appears

that males have a longer average sentence length than females. Figure 3-3 shows the average document length. The average document length is defined as the number of tokens (Table A-1) divided by the number of documents (Table 3-1). Appendix A contains the table of actual values of the points on this graph. Similar to the previous graph, there is a peak for the 30-39 age group. However, it is interesting to note that on average, females tend to have more tokens per document than males. These two graphs suggest that females tend to type more often in shorter phrases, while men tend to type less often in longer phrases.

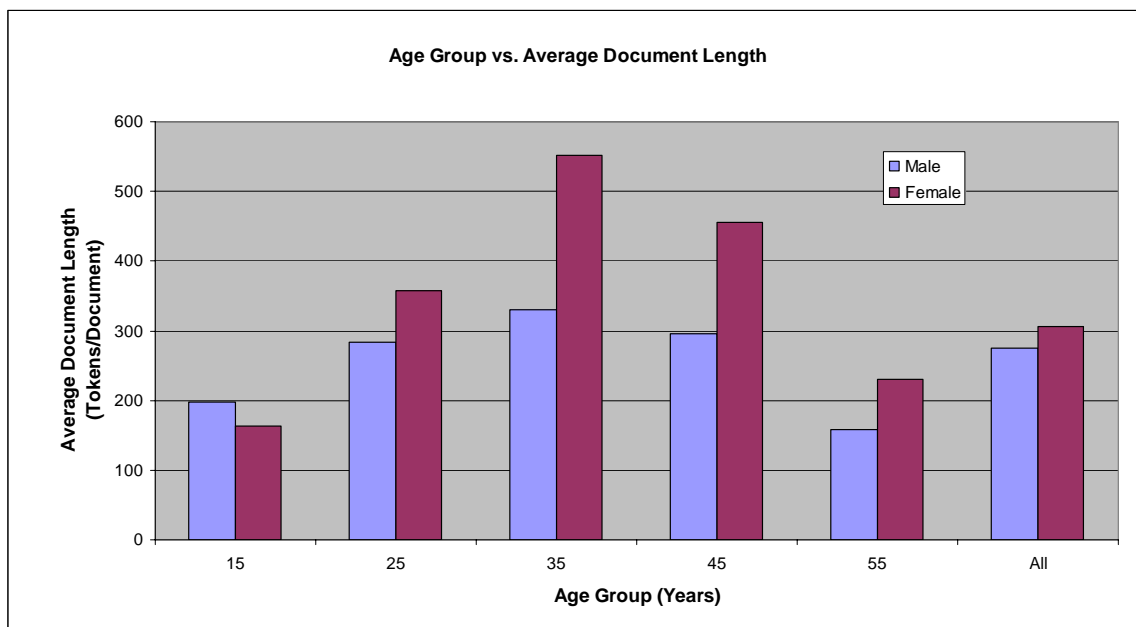


Figure 3-3. Age Group vs. Average Document Length

By decomposing Figure 3-2, we get Figures 3-4 and 3-5 which show the plot of the sentence length for each male and female document, respectively. Thus, each point on these graphs represents exactly one document. Although the linear regression lines on both graphs seem to indicate that the sentence length decreases with the increase of

age, it is very different from the picture depicted in Figure 3-2. Appendix B gives more ways of grouping the age groups by changing the binning size, including just averaging each age interval to produce one point.

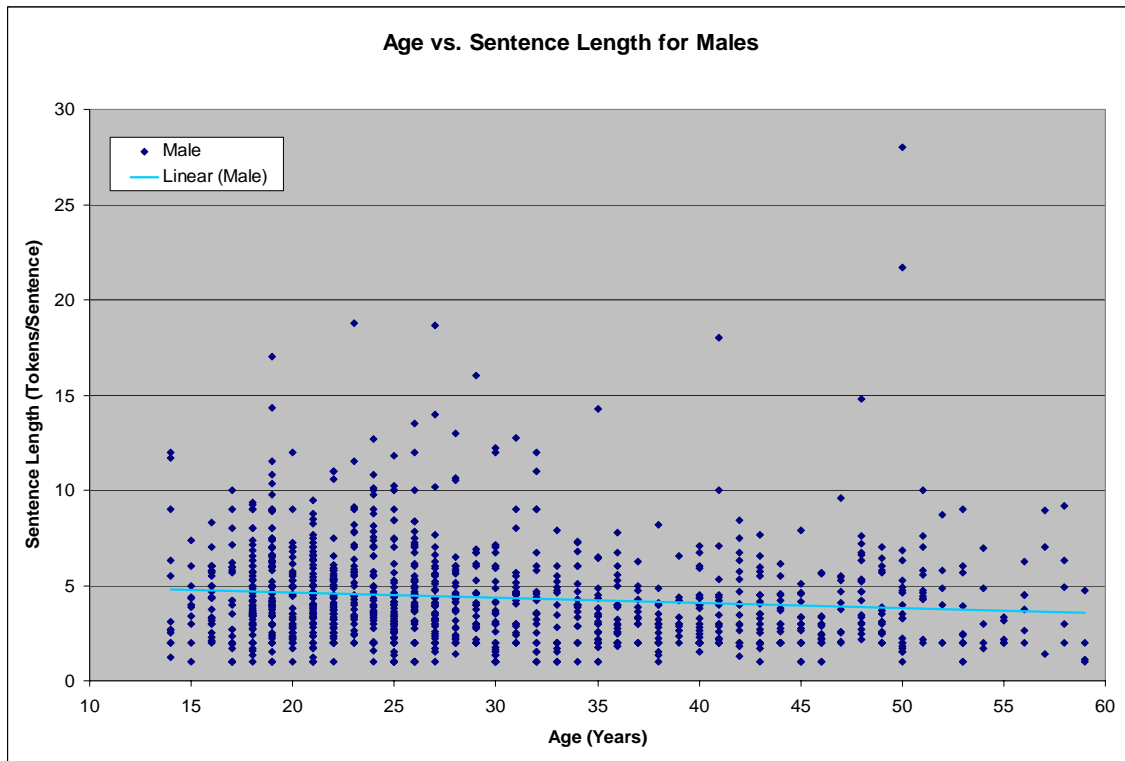


Figure 3-4. Age vs. Sentence Length for Males

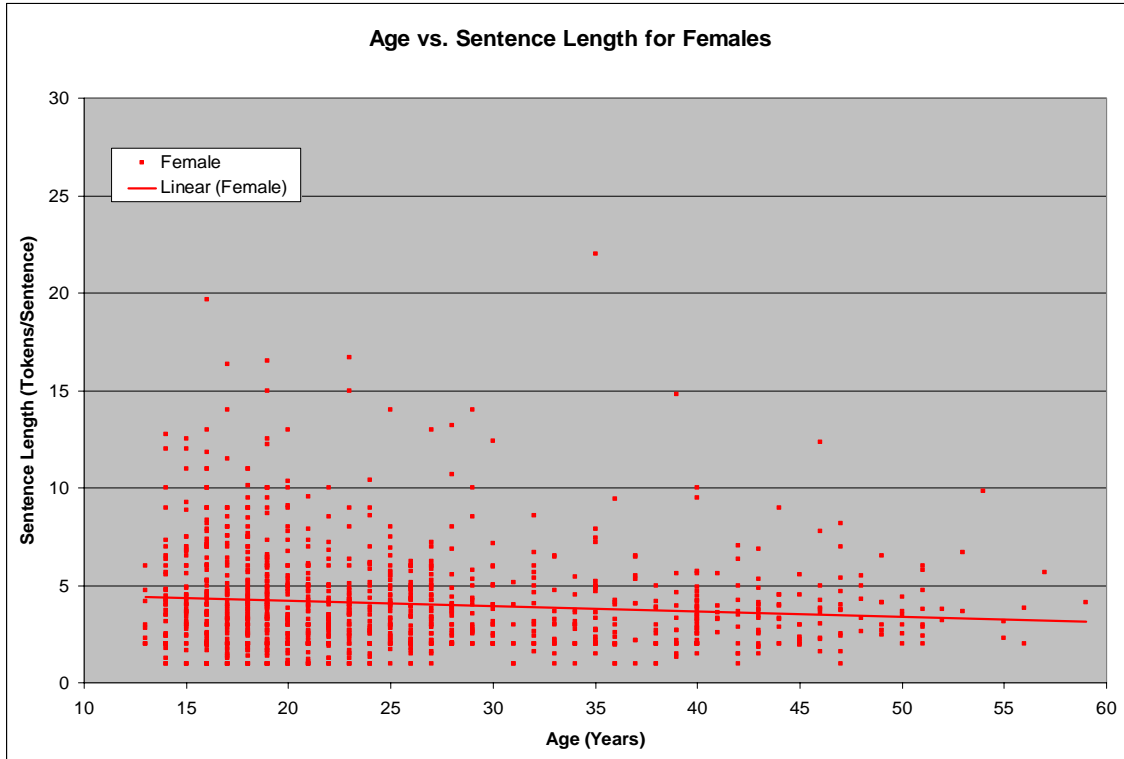


Figure 3-5. Age vs. Sentence Length for Females

Similarly, Figures 3-6 and 3-7 are zoomed in versions of the decomposed graph of Figure 3-3. The full graph is included in Appendix B. While the regression line shows a positive slope relationship between the document length and the age of females, it is fairly flat for men.

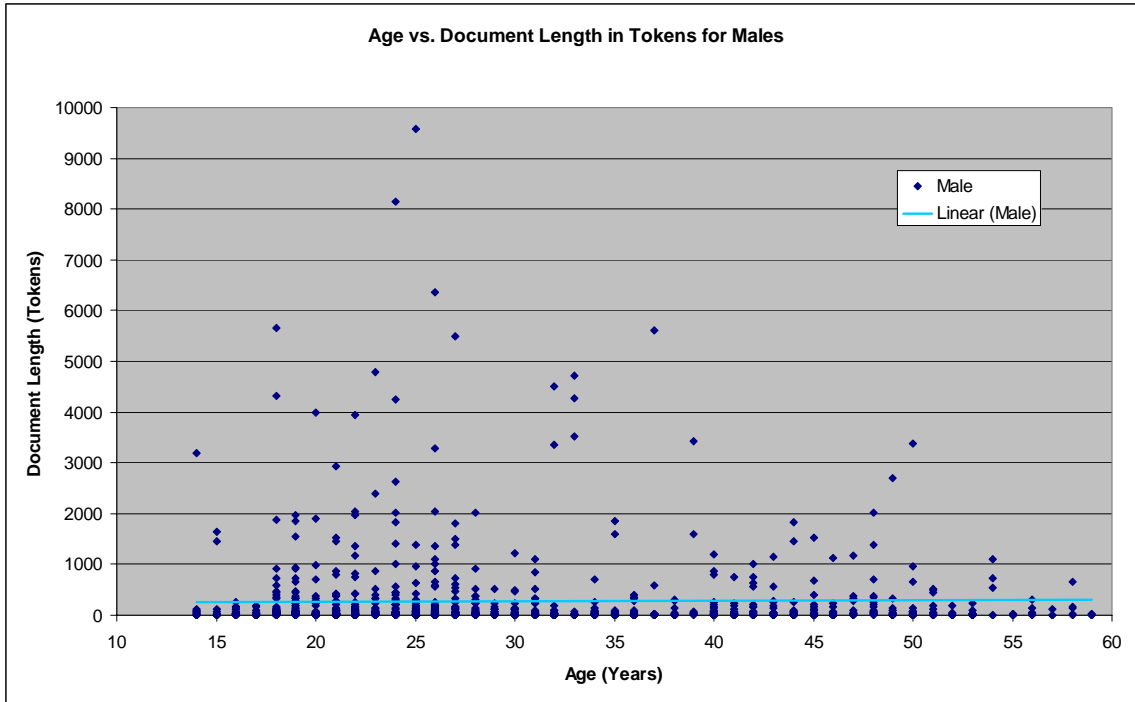


Figure 3-6. Age vs. Document Length in Tokens for Males

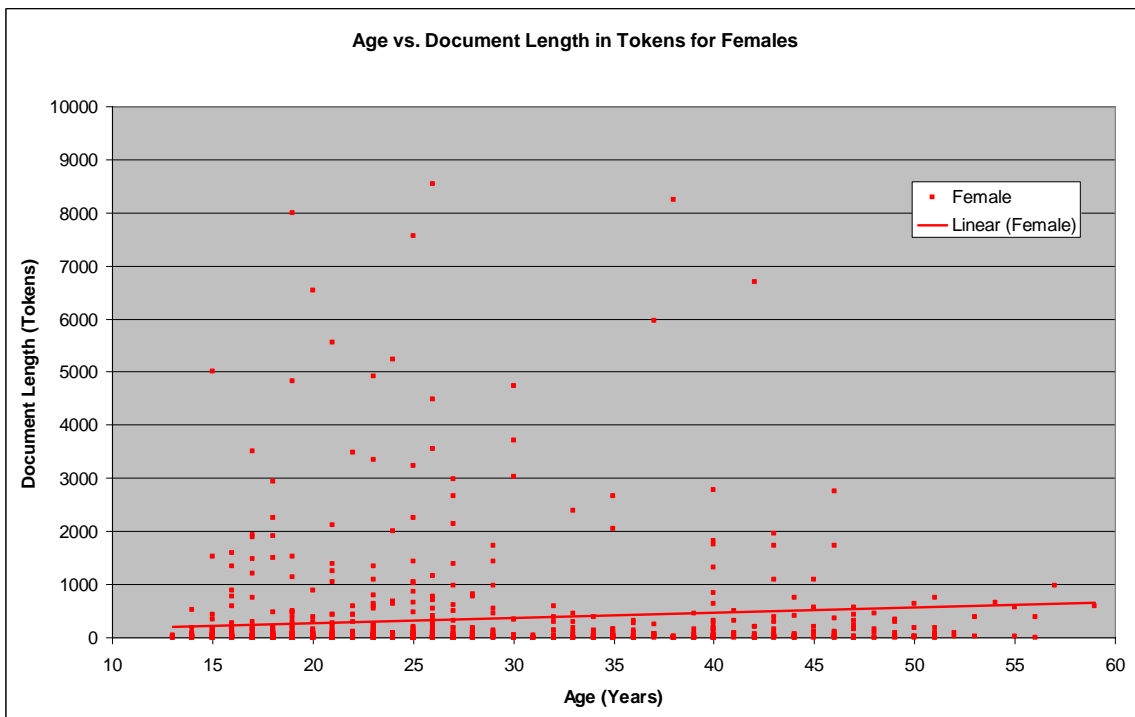


Figure 3-7. Age vs. Document Length in Tokens for Females

2. Vocabulary Variety

Figure 3-8 shows a summary of the variety in vocabulary for both genders grouped as follows: 13-19 is labeled as 15, 20-29 is labeled as 25, 30-39 is labeled as 35, 40-49 is labeled as 45, and 50-59 is labeled as 55. Agreeing with Singh's study mentioned in Section II-B, men appear to on average have a higher vocabulary range than females with the last age group fitting the exception. Vocabulary variety is measured by counting the number of types or unique tokens (Appendix C) and dividing that by the number of tokens (Appendix A). Appendix C contains the table of actual values of the points on this graph.

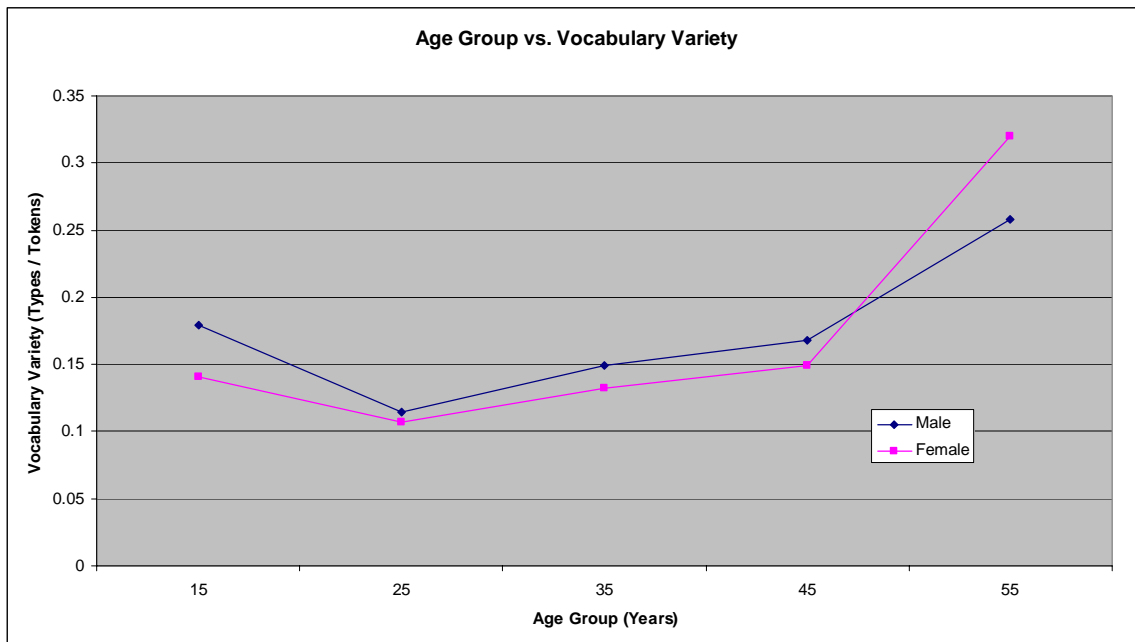


Figure 3-8. Age Group vs. Vocabulary Variety

3. Emoticons

Built-in emoticons of the chat system include the following:

Emoticon	Icon Displayed
:-(Blue Frowning Face
:-)	Yellow Smiley Face
:-@	Red Angry Face
:-o	Yellow Surprised Face
;-)	Yellow Winking Face
>:->	Yellow Mischievous Face
:beer:	Brown Beer Mug
:blush:	Yellow Blushing Face
:love:	Two Pink Hearts
:tongue:	Yellow Face with Tongue

Table 3-4. Built-In Emoticons

Figure 3-9 and 3-10 show a summary of the average emoticon token per sentence for males and females, respectively. This is calculated by counting the total number of built-in emoticons (Appendix D) and dividing it by the total number of sentences (Appendix A) per emoticon per gender. The age group is again grouped with ages 13-19 labeled as 15, ages 20-29 labeled as 25, ages 30-39 labeled as 35, ages 40-49 labeled as 45, ages 50-59 labeled as 55.

It is interesting to see the differences in emoticon frequencies between the two genders. For example, males between the ages of 50 and 59 use a lot more :love: emoticons than females of the same age who use more :-) emoticons.

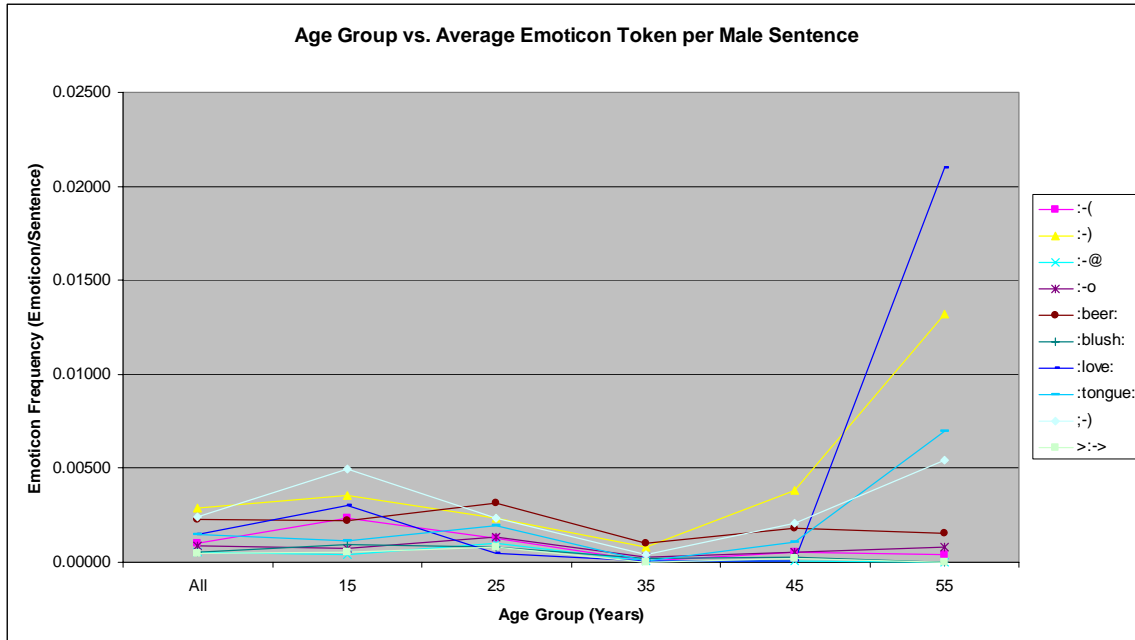


Figure 3-9. Age Group vs. Average Emoticon Token per Male Sentence.

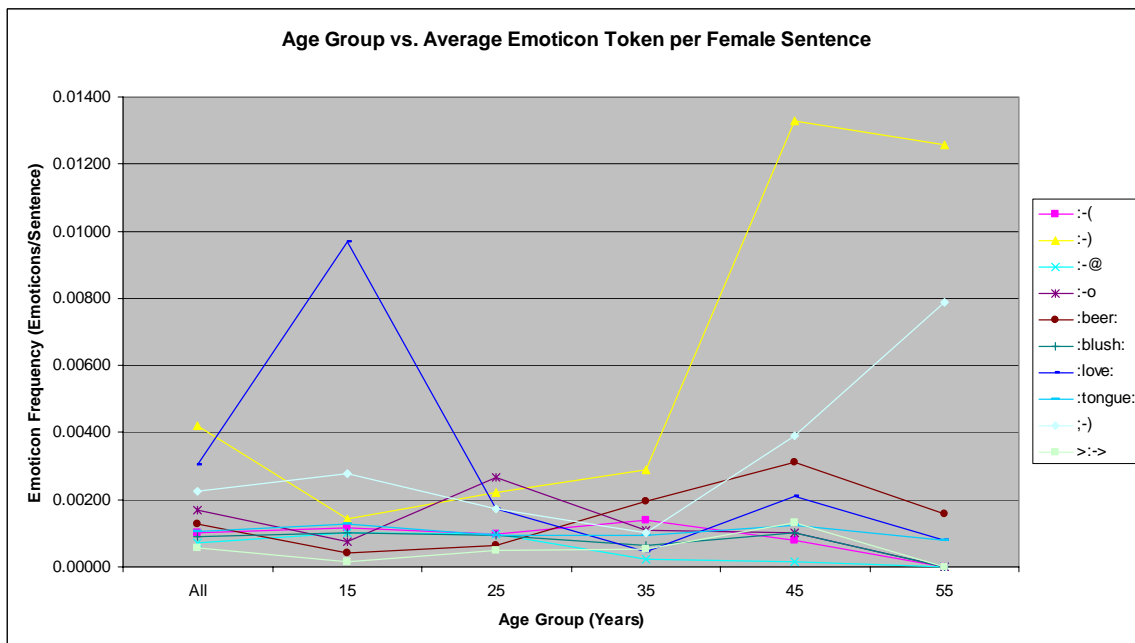


Figure 3-10. Age Group vs. Average Emoticon Token per Female Sentence

Appendix D also includes a graph of the average emoticon token with both genders combined. For emoticon

types, we counted the number of sentences with a specific emoticon and divided it by the total number of sentences. Appendix D has more tables and graphs of emoticon types as well.

4. Punctuations

In terms of punctuation marks, we included all possible punctuation marks on a standard QWERTY keyboard. Figures 3-11 and 3-12 show the average punctuation tokens for male sentences and female sentences, respectively. To calculate the average punctuation tokens, we divided the total number of a specific punctuation mark by the total number of sentences per gender.

It is important to comment on the popularity of the period (.). In this particular chat host, ascii .'s are used in system messages and/or settings as well. For example, when a user wants to express some action, it is represented as ".ACTION" in the chat log. Another example is using a period followed by a specific number to represent font color or font size. Many other system messages exist that contribute to the high count of periods. One reason for contributing to high counts of ('s and)'s is because a series of these represents a hug. (((Molly))) would mean that Molly is getting a hug. The number of open and closed parenthesis used to signify a hug also varies from two or to 10 or more. Besides these three punctuation marks, there does not appear to be that much difference in terms of gender.

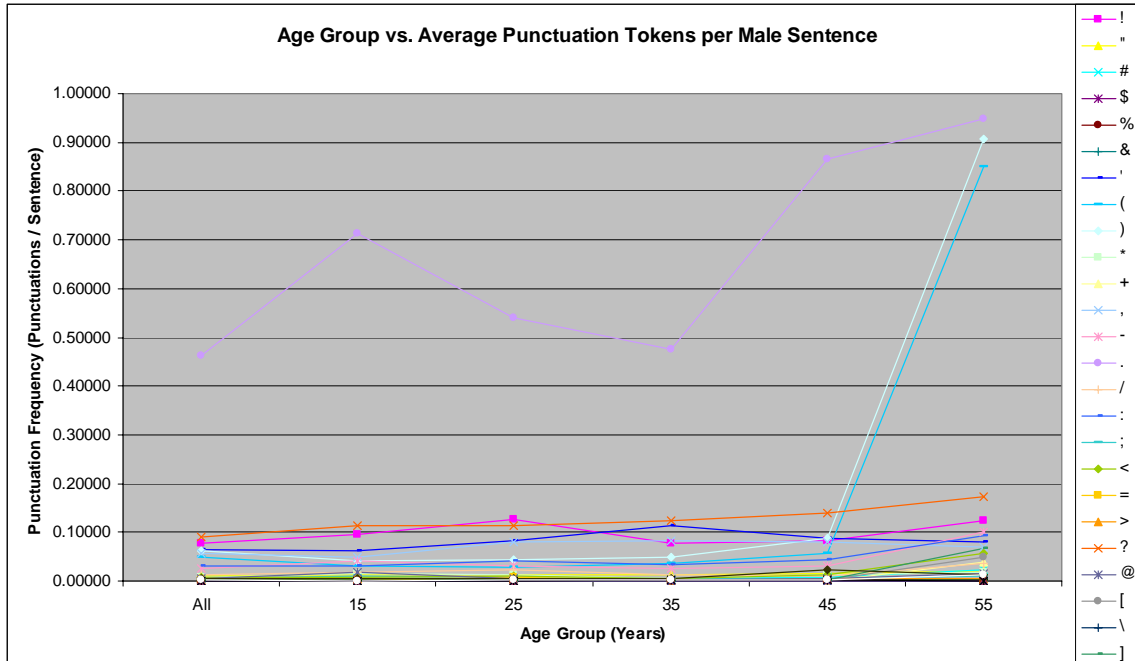


Figure 3-11. Age Group vs. Average Punctuation Tokens per Male Sentence

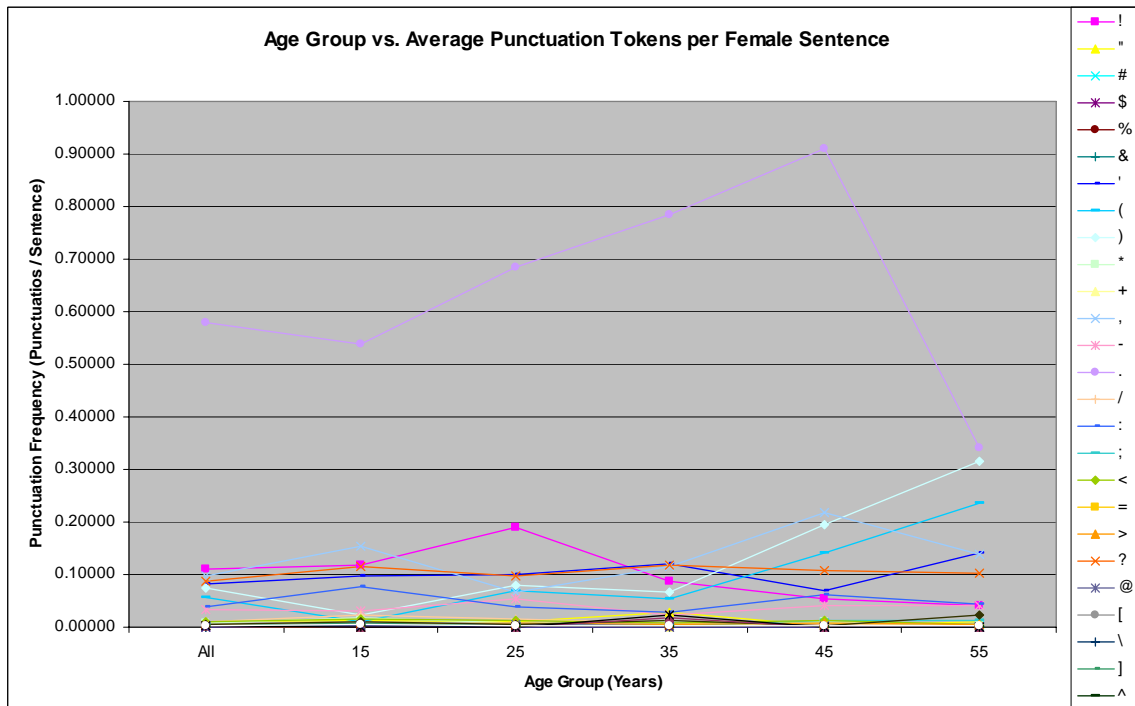


Figure 3-12. Age Group vs. Average Punctuation Tokens per Female Sentence

Appendix E also includes a graph of the average punctuation token with both genders combined. For punctuation types, we counted up the number of sentences with a specific punctuation and divided it by the total number of sentences. Appendix E has more tables and graphs of punctuation types as well.

D. CHAPTER SUMMARY

In this chapter, the generation of the corpus is discussed. The description of the data is provided. The major part of this chapter contained statistical analysis mainly on the training set portion of the data. There appear to be differences in age and gender in some of the features that may be useful for building automatic authorship attribution systems. The next chapters includes the profiling of chat logs and the machine learning and classification tools used.

IV. TESTING AND ANALYSIS

A. MACHINE LEARNING AND CLASSIFICATION TECHNIQUES

Machine learning allows the adaptation to new circumstances and the detection and extrapolation of patterns from seen circumstances [26]. While there are many machine learning tools available, this research only involved the Naïve Bayes methodology.

1. Classification Tool

Not only is the Naïve Bayesian Classifier one of the most effective classification algorithms, it is probably also the most common Bayesian network model [26,36]. Although it makes independence assumptions, it may also be a good classifier when the features are not completely independent [26].

Naïve Bayes is a simple probabilistic classifier based on applying Bayes' theorem using independence assumptions. These independence assumptions are what make it *naïve*. The probability model for a classifier is a conditional model. Assume that we have a list of features (F_1, \dots, F_n) and some class (C) , the probability of which is conditional on this set of features. Combined with Bayes' theorem, we get:

$$P(C | F_1, \dots, F_n) = \frac{P(C)P(F_1, \dots, F_n | C)}{P(F_1, \dots, F_n)}$$

Since we are interested in making a classification decision and the denominator does not depend on C , we can continue by ignoring it completely. Thus, by assuming independence amongst the features, the numerator can now be simplified and our model becomes:

$$P(C | F_1, \dots, F_n) \propto P(C) \prod_{i=1}^n P(F_i | C)$$

Taking the argmax_C over all different C 's will give the most probable class given those set of features.

2. Classification Method

Naïve Bayes provides the ability to choose the class with the highest probability to serve as the label. In this research, we came across two different situations - binary and multi-class classification.

In a binary classification, there are only two distinct choices for labels. For example, the gender is either male or female and cannot be both. We have also grouped ages together to create a binary classification problem of teens (13-19) or 20's (20-29), teens or 30's (30-39), teens or 40's (40-49), teens or 50's (50-59), and so on with every binary combination between the five age groups. Following up on the finding that 86% of the offenders are 26 or older, we also looked at the binary classification of 13-15 or 26-59 (Table 1-1).

The only multi-class classification problem was that of classifying all five age groups at once. In this case, the class with the overall highest probability was chosen as the label.

3. Measures of Classification Performance

Before calculating classification performance, it is necessary to assign the result of each classification to one of the following four result types:

- True Positive (TP) - the classifier identified a positive class data point as positive;

- False Negative (*FN*) - the classifier has identified a positive class data point as negative;
- False Positive (*FP*) - the classifier has identified a negative class data point as positive;
- True Negative (*TN*) - the classifier has identified a negative class data point as negative.

In looking at gender, an actual male that is labeled male makes a True Positive; an actual male that is labeled female makes a False Negative; an actual female that is labeled a male makes a False Positive; and actual female that is labeled a female makes a True Negative. For all binary classifications (e.g. gender and age), a two-way confusion matrix can be constructed as shown in Figure 4-1.

Predicted Class			
Actual Class		Yes	No
	Yes	True Positive	False Negative
	No	False Positive	True Negative

Figure 4-1. Two-Way Confusion Matrix

For multi-class classifications of age, a slight modification is used to create a Five-Way Confusion Matrix as shown in Figure 4-2. The names of the result types have been modified for clarity purposes.

		Predicted Class				
Actual Class		13-19	20-29	30-39	40-49	50-59
	13-19	True Teens	False 20s	False 30s	False 40s	False 50s
	20-29	False Teens	True 20s	False 30s	False 40s	False 50s
	30-39	False Teens	False 20s	True 30s	False 40s	False 50s
	40-49	False Teens	False 20s	False 30s	True 40s	False 50s
	50-59	False Teens	False 20s	False 30s	False 40s	True 50s

Figure 4-2. Five-Way Confusion Matrix

Given the confusion matrices, various measures of evaluating classification performance are now available. The measures used in this study include:

- Precision (P)
- Recall (R)
- F-score (F)

Precision measures the proportion of objects in the result set that are actually relevant, and recall measures the proportion of all the relevant objects in the collection that are in the result set [26]. For binary classification problems, they are defined as follows:

$$\text{Precision } (P) = \frac{TP}{TP+FP}$$

$$\text{Recall } (R) = \frac{TP}{TP+FN}$$

A system can easily trade off precision for recall. For example, a system that returns the same label for every input will have a recall score of 100%, but a poor precision score. More precisely, the precision score will equal the proportion of the input that actually is the label. Thus, it is important to have a good balance between the precision and the recall. The F-Score is a version of the harmonic mean and serves this purpose by combining the precision and recall values into a single value:

$$\text{F-Score } (F) = \frac{2}{1/P + 1/R}$$

$$\text{Harmonic Mean} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

While two precision, two recall, and two f-scores can be calculated for binary classifications, we need a slight modification for the multi-class classification problems. To calculate the first precision (P_1) in Figure 4-1, the true positive is the value in the "true teens" box, while the false positive is the sum of all the "false teens" boxes. To calculate the first recall (R_1), the true positive remains the same as in P_1 , and the false negative is the sum of "false 20s," "false 30s", "false 40s", and "false 50s" in the 13-19 row. With P_1 and R_1 , we can

calculate the first f-score (F_1). The same procedure is repeated for the rest of the columns in Figure 4-1.

B. FEATURES AND FEATURE VECTORS

In this study, we look at 6 feature vectors comprised of individual features. A token is defined as a unit of occurrence. A type is defined as a unique unit of occurrence. The 6 feature vectors are described below.

- Emoticon Tokens - Measure of emoticon usage in the document. Measured by counting the total number of appearances of a specific emoticon.
- Emoticon Types per Sentence - Measure of unique emoticons in a sentence. Measured by counting the total number of sentences with a specific emoticon.
- Punctuation Tokens - Measure of punctuation usage in the document. Measured by counting the total number of appearances of a specific punctuation.
- Punctuation Types per Sentence - Measure of unique punctuations in a sentence. Measured by counting the total number of sentences with a specific punctuation.
- Word Tokens - Measure of average sentence length in the document. Measured by counting the total number of tokens and dividing it by the total number of sentences.
- Word Types - Measure of vocabulary variety in the document. Measured by counting the total number of type and dividing it by the total number of tokens in the document.

Within each feature vector, we have a varied number of features. The emoticon feature vectors each consist of 8 features that are comprised of built-in emoticons illustrated in Table 3-4. The punctuation feature vectors each consist of 32 features that are comprised of all punctuation characters found on a standard QWERTY keyboard. The word feature vectors have one feature associated with each.

C. EXPERIMENT SETUP

1. Creating the Joint Probability Distribution Tables from the Training Data

The Joint Probability Distribution Table (JPDT) is the fundamental element for the Naïve Bayes Classifier. JPDTs for Emoticon Tokens, Emoticon Types, Punctuation Tokens, Punctuation Types, Word Tokens, and Word Types all have each element in a class on one axis and bins on the other axis. The three steps involved in creating the JPDTs are:

- Step 1: Counting Tokens and Types
- Step 2: Binning into n bins
- Step 3: Smoothing with Witten-Bell Discounting

Figures 4-3 to 4-5 describe these three steps in pseudo code.

Input: files from training corpus
<p>For each input and feature in each feature vector</p> <p>For Emoticon, Punctuation, Word Tokens</p> <p>Calculate the bin for each token feature by the taking the count of the number of tokens (e.g. :-) or !) and dividing it by the total number of sentences in the document; Increment the count for this bin and feature</p> <p>For Emoticon, Punctuation Types</p> <p>Calculate the bin for each type feature by taking the number of sentences with the types (e.g. :-) or !) and dividing it by the total number of sentences in the document; Increment the count for this bin and feature</p> <p>For Word Types</p> <p>Calculate the bin by taking the number of word types and dividing it by the number of tokens in the document; Increment the count for this bin and feature</p>
<p>Output: 82 counts (8 emoticon tokens, 8 emoticon types, 32 punctuation tokens, 32 punctuation types, 1 word token, 1 word type) in a bin per file</p>

Figure 4-3.

Step 1 - Counting Tokens and Types

Input: Output from Step 1
<p>For each feature in each feature vector</p> <p>Find the smallest bin (bin_{Small}) and the biggest bin (bin_{Big})</p> <p>Calculate the bin size by: $(bin_{Big} - bin_S) / n$</p> <p>Redistribute the raw counts into the appropriate bin (all original bins (bin_{orig}) will fall into one of the n bins by: $(int(bin_{orig} - bin_{Small})) / n$</p>
<p>Output: n bins for each feature of each feature vector where each table element is still the raw document count</p>

Figure 4-4.

Step 2 - Binning into n bins

Input: Output from Step 2
<p>For each feature in each feature vector</p> <p> For each classification in each class (e.g. male or female in gender or teens, 20s, 30s, 40s, or 50s in age)</p> <p> Calculate N (Number of documents) by: summing up of all the raw counts in the table</p> <p> Calculate T (Number of different bins) by: (<i>number of classifications per class</i>) * (<i>NUMBER OF BINS</i>)</p> <p> Calculate Z (Number of bins with a zero-count)</p> <p> For all zero-count bins</p> <p> Apply the formula: $T / (Z * (N + T))$</p> <p> For all non-zero bin</p> <p> Apply the formula: $Count / (N + T)$</p>
Output: JPDTs with no non-zero elements in any of the JPDTs and n bins for each feature in each feature vectors

Figure 4-5. Step 3 – Smoothing with Witten-Bell Discounting [17]

2. Labeling the Test Data

After the JPDTs are created, we can now label our test data with the following two steps:

- Step 1: Calculating the Bins
- Step 2: Finding the Best Label for Feature Vectors with Naïve Bayes

As aforementioned, the formula we are using is:

$$P(C | F_1, \dots, F_n) \propto P(C) \prod_{i=1}^n P(F_i | C)$$

Following the description of these two steps in Figures 4-6 and 4-7, we present two alternate Step 2s.

Input: files from testing corpus
<p>For each input and feature in each feature vector</p> <p>For Emoticon, Punctuation, Word Tokens</p> <p>Calculate the bin for each token feature by the taking the count of the number of tokens (e.g. :-) or !) and dividing it by the total number of sentences in the document</p> <p>For Emoticon, Punctuation Types</p> <p>Calculate the bin for each type feature by taking the number of sentences with the types (e.g. :-) or !) and dividing it by the total number of sentences in the document</p> <p>For Word Types</p> <p>Calculate the bin by taking the number of word types and dividing it by the number of tokens in the document</p> <p>If any of the bins is bigger than the bin_{Big}, then use bin_{Big}; if any of the bins is smaller than bin_{Small}, then use bin_{Small}</p> <p>Take the original (adjusted) bin, and calculate which bin it falls into by; $(int(bin_{Orig} - bin_{Small})) / n$</p>
Output: 82 bin calculations (8 emoticon tokens, 8 emoticon types, 32 punctuation tokens, 32 punctuation types, 1 word token, 1 word type) per file

Figure 4-6. Step 1 - Calculating the Bins

<p>Input: JPDTs from the training data, output from Step 1</p> <p>For each classification (C_k) in each class (e.g. male or female in gender or teens, 20s, 30s, 40s, or 50s in age) and each Feature Vector F</p> <p>Look up the $P_i(\text{Bin} \wedge C_k)$ in the JPDT for each feature F_{ij} in feature vector F_i</p> <p>Calculate the conditional probability $P(\text{Bin} C_k)$ for each F_{ij}: Divide the joint probability by the marginal probability $P(C_k)$ where $P(C_k)$ is calculated by summing up the joint probabilities in all the bins for C_k</p> <p>Multiply all the conditional probabilities $P(\text{Bin} C_k)$ together and multiple this by $P(C_k)$ to complete the Naïve Bayes</p> <p>For Emoticon, Punctuation Types</p> <p>Calculate the bin for each type feature by taking the number of sentences with the types (e.g. :-) or !) and dividing it by the total number of sentences in the document</p> <p>For Word Types</p> <p>Calculate the bin by taking the number of word types and dividing it by the number of tokens in the document</p> <p>Output: For each class, $P(C_1 F_{i1}, \dots, F_{in})$, $P(C_2 F_{i1}, \dots, F_{in})$, ..., $P(C_m F_{i1}, \dots, F_{in})$ where m is the number of classifications in each class and label C_x for which $P(C_x F_1, \dots, F_n)$ is greater than all other $P(C_k F_1, \dots, F_n)$, $k \neq x$</p>

Figure 4-7. Step 2 – Finding the Best Label Using Naïve Bayes

a. Alternate Step 2: Finding the Best Label for Individual Features with Naïve Bayes

Because so many features are grouped together in each feature vector, we decided to look at the affects of each individual feature. Because this is very similar to Step 2 in Figure 4-7, except we look at individual features instead of combining all the features of each feature vector together, we can make some simple modifications to

Figure 4-7. After skipping the steps "Calculate the conditional probability..." and "Multiple all conditional probabilities..." and comparing the probability directly from the JPDT since $(P(\text{Bin}/C_k) / P(C_k)) * P(C_k) = P(\text{Bin}/C_k)$, we choose the label C_x for which $P(C_x/F_i)$ is greater than all other $P(C_k/F_1, \dots, F_n)$, $k \neq x$.

b. More Alternate Step 2's: Finding the Best Label for Feature Vectors with Naïve Bayes without the influence of the Prior

Due to the an overwhelming amount of labels being the prior ($P(C)$), we ran the same experiments again for feature vectors and individual features, but removed the influence of the prior in the step "...multiply this by $P(C_k)$ to complete the Naïve Bayes," to see if there are any noticeable changes in the resulting f-scores. Results from these experiments are detailed in the next section.

D. RESULTS AND ANALYSIS

In this research, we looked at labels for each feature vector as well as each feature for both gender and age. From Appendix F and G, we extracted features and feature vectors that have f-scores higher than the baseline f-scores or precisions higher than $\max(0.8, \text{baseline precision})$. In the next two sections, we see that there are a number of features and feature vectors that meet this requirement.

Higher f-scores let us guess the class better than simply guessing based on the distribution of classes from the training data (i.e. the prior). However, since the f-scores are not significantly higher than the baseline f-scores shown in Appendix F and G, we cannot conclude that individual features or feature vectors using Naïve Bayes is a good indicator of class. Nevertheless, there is still

hope that these features and feature vectors will be able to distinguish ages and gender when looked at from a different perspective. This is elaborated in Chapter V.

Although f-scores in this experiment did not produce significantly good results, we did have better luck with the precision measure. Higher precision lets us be more confident in that the class labeled is more likely to be the actual class. This is extremely useful when we want to be relatively confident in picking out only female documents from a set of unlabeled data (i.e. gender information is not available with the document).

1. Including the Influence of the Prior

Tables 4-1 to 4-9 show the summary of Feature Vectors (FV) and features which have a higher f-score than the baseline f-score or a higher precision than $\max(0.8, \text{baseline precision})$. The tables are organized by classes (male, female, teens, 20s, 30s, 40s, 50s, Under 26, 26 and Over). The keys for the feature vectors and features are included in Appendix H. An "x" means that there are no feature vectors or features that have higher f-scores than the baseline f-scores or higher precisions than the $\max(0.8, \text{baseline precision})$. "Context" means which experiment the FV and features were extracted from. For example, "All" means that data across all ages were included and "Teens/20s" means that only teens (13-19) and 20s (20-29) were included in the experiment. The tables for all the raw precision, recall, and f-scores for all "Contexts" are in Appendix F and G.

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
All	1, 2	11, 15, 16, 18, 20, 25, 26, 38, 47, 69, 81	Teens/20s	x	18, 28, 31, 32, 42, 53, 71, 73
20s/30s	x	11, 15, 20, 36, 49, 71	Teens/30s	x	1, 3, 4, 35, 42
20s/40s	x	2, 15, 16, 36	Teens/40s	x	11, 39, 42, 59, 68, 71
20s/50s	x	2, 11, 15	Teens/50s	x	42, 52, 71
30s/40s	1	4, 45			
30s/50s	x	2, 47, 48, 49, 83			
40s/50s	1, 2	4, 45, 46			
26OrNot	1	9, 16, 18, 20, 26, 35, 37, 38, 47, 56			

Table 4-1. Extracted Features and Feature Vectors for Males Including the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
Teens/20s	x	11, 18, 28, 31, 32, 42, 53, 71, 73	All	1	15, 20, 25, 26, 38, 47, 69, 81
Teens/30s	x	1, 3, 4, 16, 35, 42	20s/30s	x	11, 15, 20, 36, 49, 71
Teens/40s	x	11, 16, 39, 42, 59, 68, 71	20s/40s	x	2, 15, 16, 36
Teens/50s	x	11, 12, 16, 42, 52, 71	20s/50s	x	2, 15
			30s/40s	x	4, 45
			30s/50s	x	2, 47, 48, 49, 83
			40s/50s	1, 2	4, 45, 46
			26OrNot	1	9, 16, 18, 20, 26, 37, 38, 47, 56

Table 4-2. Extracted Features and Feature Vectors for Females Including the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
Teens/30s	x	3,19,20,50	All	x	23,26,29,34,47,57,69,70
Teens/40s	4	40,68	Teens/20s	3,5	23,29,34,35,37,38,42,47,57,63,69,70,71,77

Table 4-3. Extracted Features and Feature Vectors for Teens Including the Prior

F-Score > Baseline F-Score		
Context	FV	Features
All	1	3,4,14,23,26,29,34,35,37,38,40,42,47,56,57,63,69,70,71
Teens/20s	x	14,17,23,29,34,35,37,38,42,47,57,63,69,70,71,77
20s/30s	x	3,4,14,35,37,38,48,71
20s/40s	x	26,50

Table 4-4. Extracted Features and Feature Vectors for 20s Including the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
30s/40s	2,6	2,15,22,59,68,78	Teens/30s	x	3,19,20,50
			20s/30s	x	3,4,14,35,37,38,48,71

Table 4-5. Extracted Features and Feature Vectors for 30s Including the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
30s/40s	6	x	Teens/40s	x	40,68
40s/50s	x	18,27	20s/40s	x	26,50
			30s/40s	2	2,4,15,22,59,68

Table 4-6. Extracted Features and Feature Vectors for 40s Including the Prior

Precision > Max(0.8, Baseline Precision)		
Context	FV	Features
40s/50s	x	18,27

Table 4-7. Extracted Features and Feature Vectors for 50s Including the Prior

F-Score > Baseline F-Score		
Context	FV	Features
26OrNot	x	3,4,12,50,53

Table 4-8. Extracted Features and Feature Vectors for Under 26 Including the Prior

Precision > Max(0.8, Baseline Precision)		
Context	FV	Features
26OrNot	x	3,4,12,14,38,39,50,53

Table 4-9. Extracted Features and Feature Vectors for 26 and Over Including the Prior

2. Excluding the Influence of the Prior

Since a lot of the test data were getting labeled as the class with the higher prior $P(C)$, we were interested in seeing the effects of the prior removed. Tables 4-10 to 4-18 show the summary of FVs and features which have a higher f-score than the baseline f-score or a higher precision than $\max(0.8, \text{baseline precision})$. The tables are organized first by classes (male, female, teens, 20s, 30s, 40s, 50s, Under 26, 26 and Over). The keys for the feature vectors (FV) and features are included in Appendix H. An "x" means that there are no feature vectors or features that have higher f-scores than the baseline f-scores of higher precisions than the $\max(0.8, \text{baseline precision})$. "Context" means which experiment the FV and features were extracted from. For example, "All" means that data across all ages were included and "Teens/20s" means that only

teens (13-19) and 20s (20-29) were included in the experiment. The tables for all the raw precision, recall, and f-scores for all "Contexts" are in Appendix F and G.

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
All	1	25, 26, 27, 29, 33, 38, 47, 69, 79, 81	All	x	53
Teens/20s	x	11, 18, 56	Teens/20s	x	28, 53, 71, 73
Teens/30s	x	2, 10, 17, 18, 34, 35, 46, 47, 55, 56	Teens/30s	x	1, 3, 4
Teens/40s	x	2, 3, 4, 14, 16, 17, 18, 55, 56	Teens/40s	x	11, 39, 42, 59, 68, 71
Teens/50s	x	2, 9, 10, 14, 16, 34, 56	Teens/50s	4	42, 52, 71
20s/30s	x	20, 36, 48, 49, 50, 71	20s/30s	x	14, 31
20s/40s	x	36	20s/40s	1	24, 28, 55, 68, 70
20s/50s	x	11, 56	20s/50s	x	24, 28, 50
30s/40s	1	19, 20, 45, 46, 71	30s/40s	x	15, 16
30s/50s	x	2, 19, 20, 47, 48, 50, 71, 83	30s/50s	x	1, 15
40s/50s	1	3, 4, 6, 45, 46	40s/50s	x	62
26OrNot	1	9, 18, 35, 50, 56	26OrNot	x	17, 28, 33, 39, 40

Table 4-10. Extracted Features and Feature Vectors for Males Excluding the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
Teens/20s	x	28,53,71,73	All	x	22,25,26,27,28,29,33,38,47,53,56,64,69,78,79,81
Teens/30s	x	1,3,4,16	Teens/20s	x	11,18,56
Teens/40s	x	11,39,42,59,68,71	Teens/30s	x	2,10,17,18,34,35,46,47
Teens/50s	x	11,12,42,52,71	Teens/40s	x	3,4,14,17,18,34,35,55,56
20s/30s	1	14,31,61	Teens/50s	x	2,9,10,14,34,56
20s/40s	x	24,28,55,62,68,70	20s/30s	x	20,36,48,49,50,71
20s/50s	x	24,28,50	20s/40s	x	36
30s/40s	x	15,16	20s/50s	x	56
30s/50s	x	1,15	30s/40s	x	19,20,45,46,71
40s/50s	x	62	30s/50s	x	2,19,20,48,50,71,83
26OrNot	x	17,28,33,39,40	40s/50s	1	3,4,6,45,46
			26OrNot	x	9,18,50,56

Table 4-11. Extracted Features and Feature Vectors for Females Excluding the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
All	x	10,11	All	2	x
Teens/20s	x	9,10,12,32	Teens/20s	x	23,29,35,37,38,42,47,56,57,69,70,71
Teens/30s	x	3,4,19,20,50,71	Teens/30s	x	2,15,16,24
Teens/40s	x	4,40,46,,59,60,68	Teens/40s	x	9,15,16,55,62,64
Teens/50s	x	12,52	Teens/50s	6	11,12,52,62,64,78

Table 4-12. Extracted Features and Feature Vectors for Teens Excluding the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
Teens/20s	1, 3	14,16,17,18,23,29,35,37,38,42,47,56,57,69,70,71	Teens/20s	x	9,32
20s/30s	3, 4	3,4,12,20,35,37,38,48,50,71	20s/30s	6	2,24,56,64,78
20s/40s	x	12,26,40,50,59,60,68	20s/40s	5	2,10,15,16,62,64,77
20s/50s	x	50	20s/50s	x	11,12,50,62,64,78

Table 4-13. Extracted Features and Feature Vectors for 20s Excluding the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
All	6		Teens/30s	x	3,4,19,20,50,71
Teens/30s	6	2,15,16,24,56,78	20s/30s	x	3,4,12,20,35,37,38,48,50,71
20s/30s	6	2,16,24,56,64,78	30s/40s	x	11,15,16,35,49,62
30s/40s	6	2,22,59,68,78	30s/50s	5,6	2,12,77,78
30s/50s	5	2,12			

Table 4-14. Extracted Features and Feature Vectors for 30s Excluding the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
All	5	x	Teens/40s	4	4,40,46,59,60,68
Teens/40s	x	9,15,16,55,62,64	20s/40s	x	12,26,40,50,59,60,68
20s/40s	5	2,10,15,16,62,64,77	20s/50s	6	x
30s/40s	6	11,15,16,35,49,62,78	30s/40s	x	2,22,59,68
40s/50s	x	12,14,18,51,52	40s/50s	x	3

Table 4-15. Extracted Features and Feature Vectors for 40s Excluding the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
All	6	x	Teens/50s	6	52
Teens/50s	x	12,62,64	20s/50s	x	50
20s/50s	x	12,50,62,64	30s/50s	x	2,12
30s/50s	x	77	40s/50s	x	12,14,18,51,52
40s/50s	x	3			

Table 4-16. Extracted Features and Feature Vectors for 50s Excluding the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
26OrNot	x	3,4,12,50,53,60,73	26OrNot	x	9,16,18,23,26,27,42,47,64

Table 4-17. Extracted Features and Feature Vectors for Under 26 Excluding the Prior

F-Score > Baseline F-Score			Precision > Max(0.8, Baseline Precision)		
Context	FV	Features	Context	FV	Features
26OrNot	1,3,4	2,9,15,16,18,23,26,27,42,47,64	26OrNot	x	3,4,50,53,60,73

Table 4-18. Extracted Features and Feature Vectors for 26 and Over Excluding the Prior

E. CHAPTER SUMMARY

This chapter begins with the machine learning and classification tool (Naïve Bayes) used in this research, and follows with a description of the focus of features and feature vectors. The algorithm to create the Joint Probability Distribution Table (JPDT) from the training corpus through counting, binning, and smoothing is described. The algorithm for counting and labeling the testing corpus is also described. Results from the

experiments are finally presented and the precision and f-scores are analyzed. In the next chapter, we present a summary of this thesis, the future goals, and last remarks.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS

A. SUMMARY

Author attribution is not a new topic. However, before this research, results of author attribution of online chat logs have not been published anywhere. This thesis attempts to profile authors of online chat logs looking at gender and age based on predictors that intuition say should make a difference.

Part I of this thesis is the corpus generation. Data was collected from a publicly available online chat host and parsed into documents containing all the lines of chat from a particular screen name. After all available age and gender information is associated with these documents, all original screen names were removed from the corpus for privacy protection of the users. No data retained in the corpus can be traced back to the original screen name or any other identifying information besides the age and gender. The data is then split into a training corpus (90% of the data) and a testing corpus (10% of the data).

Part II of this thesis is the machine learning and language modeling to try to predict gender and age. Naïve Bayes is used as a first-step analysis tool. Feature vectors identified in this study include Emoticon Tokens, Emoticon Types, Punctuation Tokens, Punctuation Types, Word Tokens (or sentence length), and Word Types (or vocabulary richness). Emoticon feature vectors include 8 different features; Punctuation feature vectors include 32 different features; Word feature vectors only include 1 feature. Appendix H has a list of the features and feature vectors.

Precision, recall, and f-scores were used to evaluate the predictors. To calculate the baseline measure for a class C, simply label all the data to be class C. The baseline precision is simply the proportion of the class C in the data set, and the baseline recall is 1. Now, the f-score can be calculated with the baseline precision and recall.

Although some of the predictors had f-scores higher than the baseline f-score, they were not significantly higher. On the other hand, many predictors had a precision score that was higher than $\max(0.8, \text{baseline precision})$. This is worth noting because it can be useful if we needed to pick out a specific class and be relatively confident that the actual class of the document really is the class the predictor chose.

After the experiments using Naïve Bayes, we noticed that a big part of the test corpus is merely labeled as the class with the highest prior determined by the proportion of the class seen in the training data. Thus, we ran the same experiments again, but removed the influence of the prior. This resulted in a more varied set of labels and slightly more predictors with better f-scores or precision values. However, the f-scores were still not significantly higher. Nevertheless, given the characteristics of the data and the results from those experiments, there is hope that other machine learning and classification tools may produce a set of useful predictors for gender and age.

B. FUTURE WORK

Since this thesis covers an unexplored area with many potential benefits, there is plenty of future work to

extend this research with hope of obtaining better models with existing or new predictors.

- Other binning techniques can be used. In this research, we created four evenly spaced bins based on the minimum and maximum found in the training data. Since the majority of documents seemed to be put in the lower bins, better results may be available if the bins were adjusted.
- Besides Naïve Bayes, many other machine learning and classification tools exist. In particular, Support Vector Machines (SVM) is a recognized as a good tool for text classification and is suitable for classification of tasks where there are a small number of data points and a large number of features [5].
- A voting scheme consisting of a series of features and/or feature vectors and be examined to see if there is a specific set of features and/or feature vectors that can be better at labeling a document.
- Bigrams or higher orders n-grams may provide for better classifiers. This research only looked at unigrams frequencies.
- Other features, including other ASCII emoticons besides the 8 built-in ones included in this research, misspelled words, or abbreviations such as LOL, etc., may be better predictors or may be able to be combined with existing predictors to be create better feature vector predictors.

- Studies in email authorship attribution show that a minimum size is required to have better results [5]. This may also be an issue with chat. Looking at file size effects or how the number of lines of chat in a document affects the accuracy of the labeling might be interesting.

C. CHAPTER SUMMARY AND CONCLUDING REMARKS

In this chapter, we presented a broad summary of the thesis and suggestions for future work.

Many applications of author profiling of online chat logs in civilian and government environments exist. Although the results from this research did not produce a set of strong predictors, it did show that a Naïve Bayes model is not that right way to approach this problem. Many more experiments and machine learning and classification tools can be tried as noted in the future works, and there is high hope that something that can distinguish age and gender of online chat logs will be discovered.

APPENDIX A: TABLES FOR RAW COUNTS, FILE SIZES, SENTENCE LENGTHS, AND DOCUMENT LENGTHS

This appendix contains the following tables: Token counts for the training Set, sentence counts for the training set, file sizes for the training, average sentence lengths for the training set, average document lengths for the training set, and the average file sizes and testing set.

A. RAW COUNTS FOR THE TRAINING SET

Age Group	Male	Female	Total
Unspecified	118918	81220	200138
13-19	41045	62441	103486
20-29	131427	149313	280740
30-39	70746	77753	148499
40-49	54033	53815	107848
50-59	12353	5742	18095
Total	428522	430284	858806

Table A-1. Token Count for Training Set

Age Group	Male	Female	Total
Unspecified	23323	16884	40207
13-19	9882	14696	24568
20-29	28162	35087	63249
30-39	13968	13799	27767
40-49	11496	12810	24306
50-59	2573	1271	3844
Total	89404	94537	183941

Table A-2. Sentence Count for Training Set

Age Group	Male	Female	Total
Unspecified	640741	428810	1069551
13-19	216611	327422	544033
20-29	684255	796052	1480307
30-39	372285	417760	790045
40-49	296040	291851	587891
50-59	72645	31828	104473
Total	2282577	2293723	4576300

Table A-3. File Size for Training Set

Age Group	Male	Female	Both
Unspecified	5.099	4.810	4.978
13-19	4.154	4.252	4.212
20-29	4.667	4.256	4.439
30-39	5.065	5.635	5.348
40-49	4.700	4.201	4.437
50-59	4.801	4.518	4.707
All Ages	4.793	4.551	4.669

Table A-4. Average Sentence Length
(Tokens/Sentence)

Age Group	Male	Female	Both
Unspecified	289.34	256.21	274.91
13-19	198.29	162.61	175.10
20-29	283.25	357.21	318.3
30-39	330.26	551.44	418.31
40-49	295.26	456.06	358.3
50-59	158.37	229.68	175.68
All Ages	275.22	306.69	290.14

Table A-5. Average Document Length
(Tokens/Documents)

B. RAW COUNTS FOR THE TESTING SET

Age Group	Male	Female	Total
Unspecified	71180	46059	117239
13-19	26867	38890	65757
20-29	76214	88551	164765
30-39	42951	47693	90644
40-49	32240	29772	62012
50-59	7911	3661	11572
Total	257363	254626	511989

Table A-6.

File Size for Testing Set

APPENDIX B: FIGURES OF SENTENCE LENGTHS AND DOCUMENT LENGTHS

This appendix contains figures of different age group bins vs. the average sentence length. Figures of individual document length separated by gender are also included.

A. MEASURES OF AGE GROUP VS. SENTENCE LENGTH

Figure B-1 is produced by taking the total number of tokens for each particular age and dividing it by the total number of sentences for each particular age. The age group ranges from age 13 to 59. Note that there are no points for 14 year old males or 58 year old females.

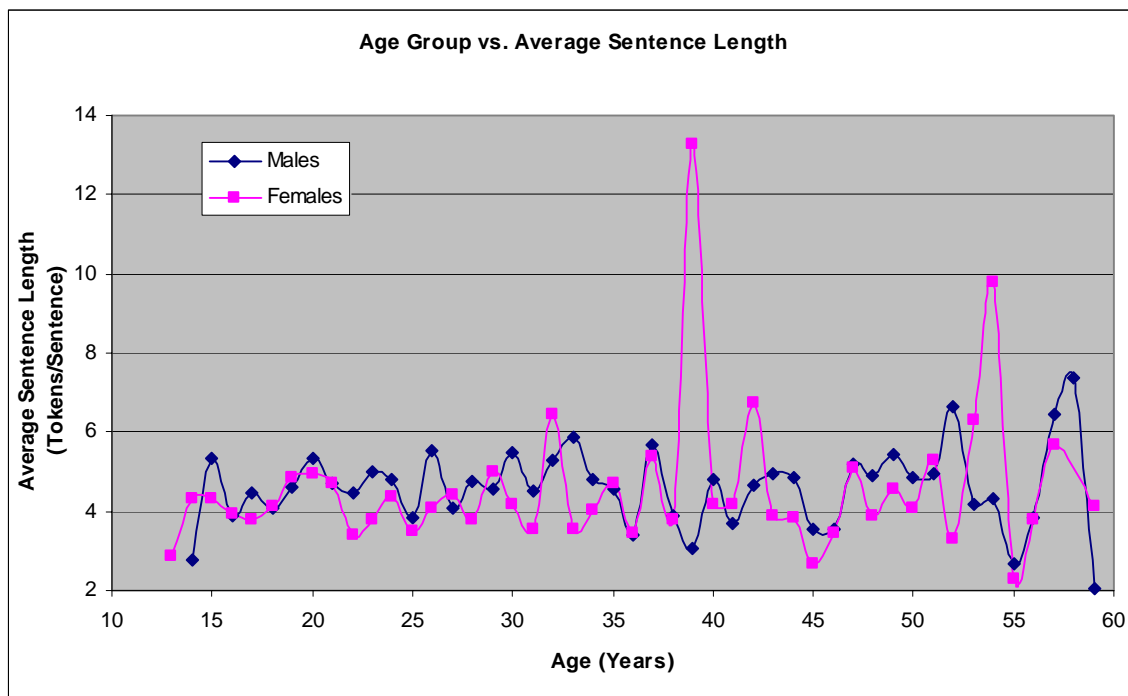


Figure B-1. Age Bin (Order 1) vs. Average Sentence Length

Figure B-2 groups the ages into bins of size 2. Thus, ages 13 and 14 form the first bin labeled as the point 13.5 on the graph, ages 15 and 15 form the second bin labeled as the point 14.5 on the graph, etc.

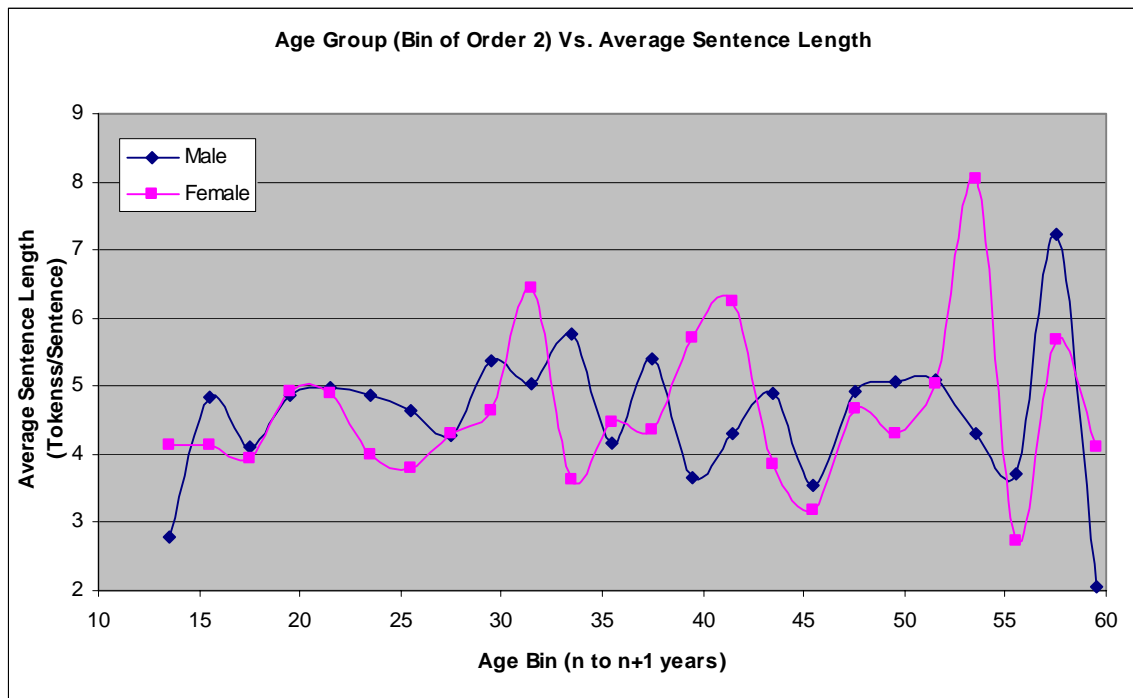


Figure B-2. Age Bin (Order 2) vs. Average Sentence Length

Figure B-3 groups the ages into bins of size 3. Thus, ages 13-15 form the first bin labeled as the point 13, ages 16-18 form the second bin labeled as the point 16, etc.

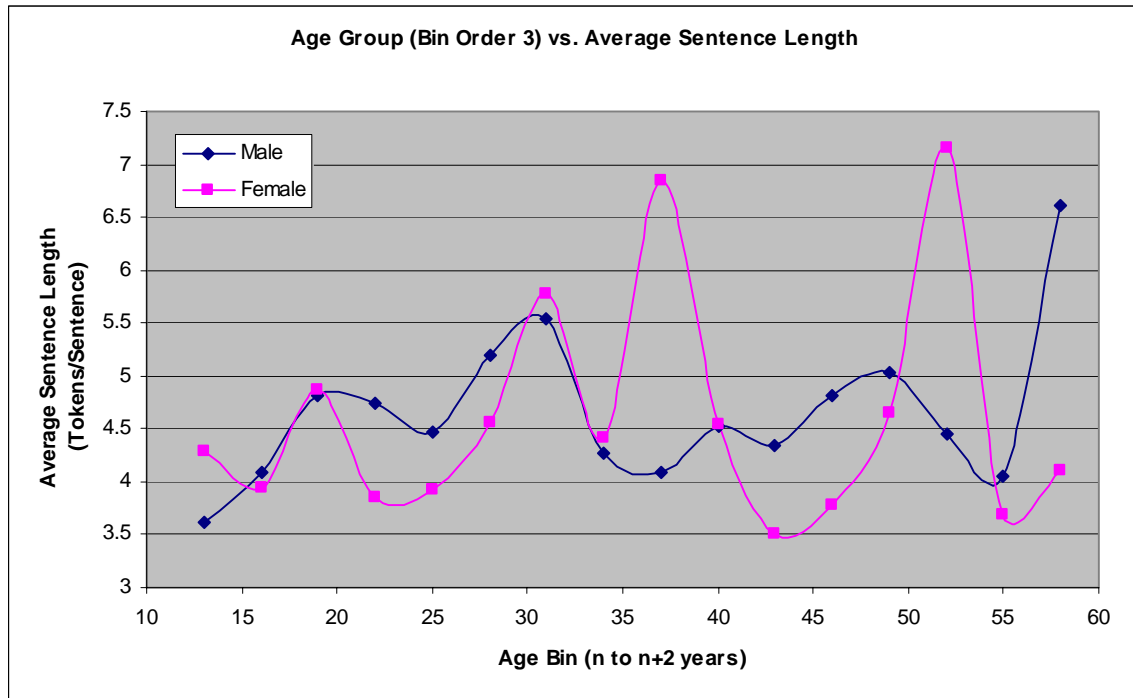


Figure B-3. Age Bin (Order 3) vs. Average Sentence Length

Figure B-4 groups the ages into bins of size 4. Thus, ages 13-16 form the first bin labeled as the point 13, ages 17-20 form the second bin labeled as the point 17, etc.

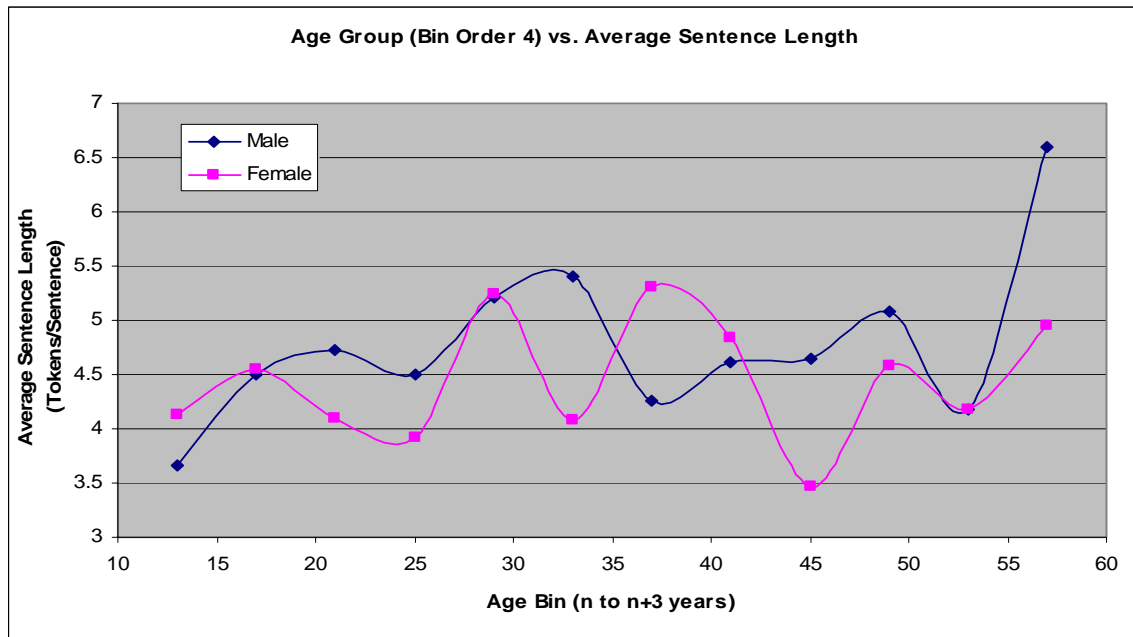


Figure B-4. Age Bin (Order 4) vs. Average Sentence Length

Figure B-5 groups the ages into bins of size 5. Thus, ages 13-17 form the first bin labeled as the point 13, ages 18-22 form the second bin labeled as the point 18, etc.

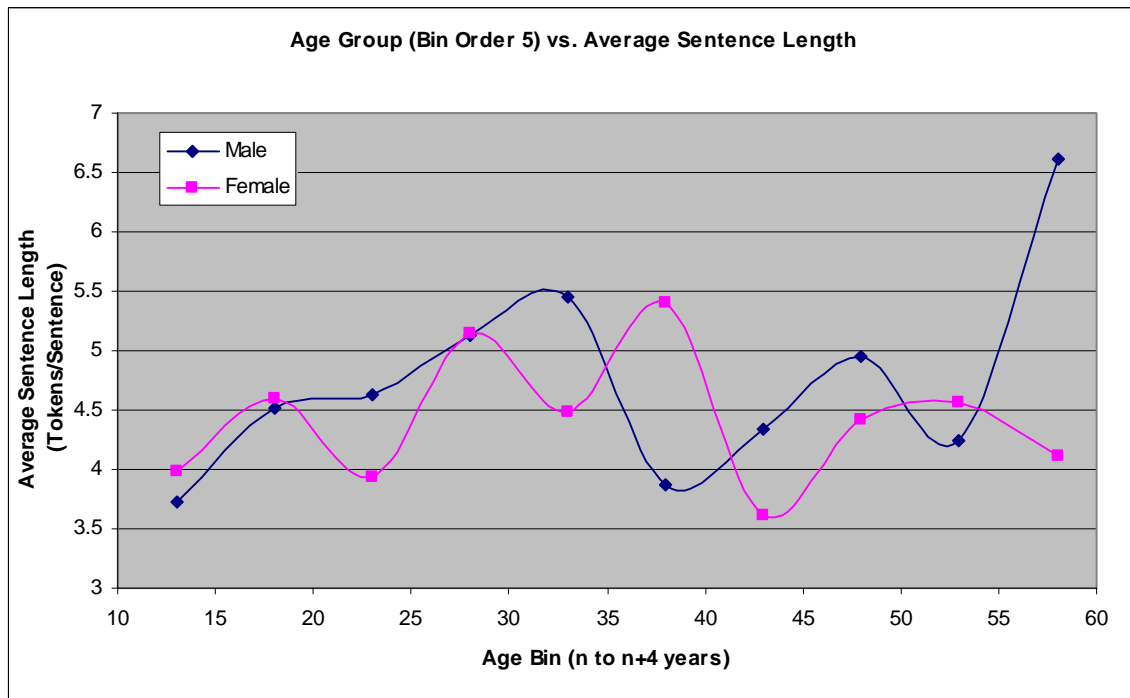


Figure B-5. Age Bin (Order 5) vs. Average Sentence Length

Figure B-6 groups the ages into bins of size 6. Thus, ages 13-18 form the first bin labeled as the point 13, ages 19-24 form the second bin labeled as the point 19, etc.

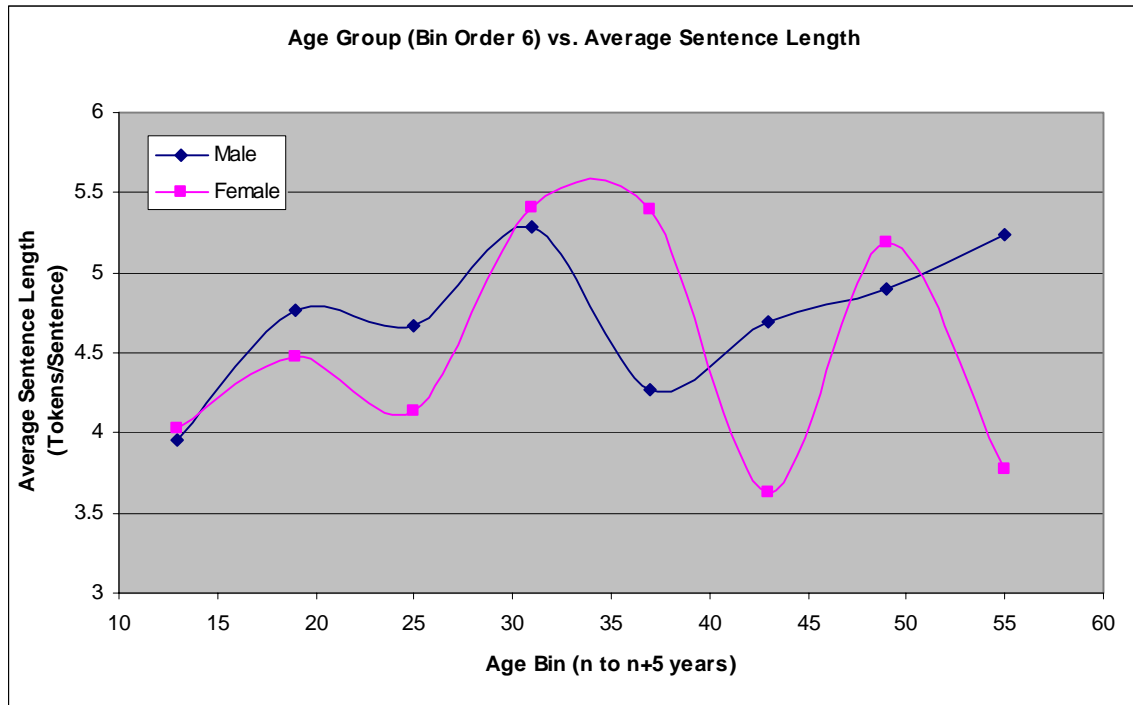


Figure B-6. Age Bin (Order 6) vs. Average Sentence Length

Figure B-7 groups the ages into bins of size 7. Thus, ages 13-19 form the first bin labeled as the point 13, ages 20-26 form the second bin labeled as the point 20, etc.

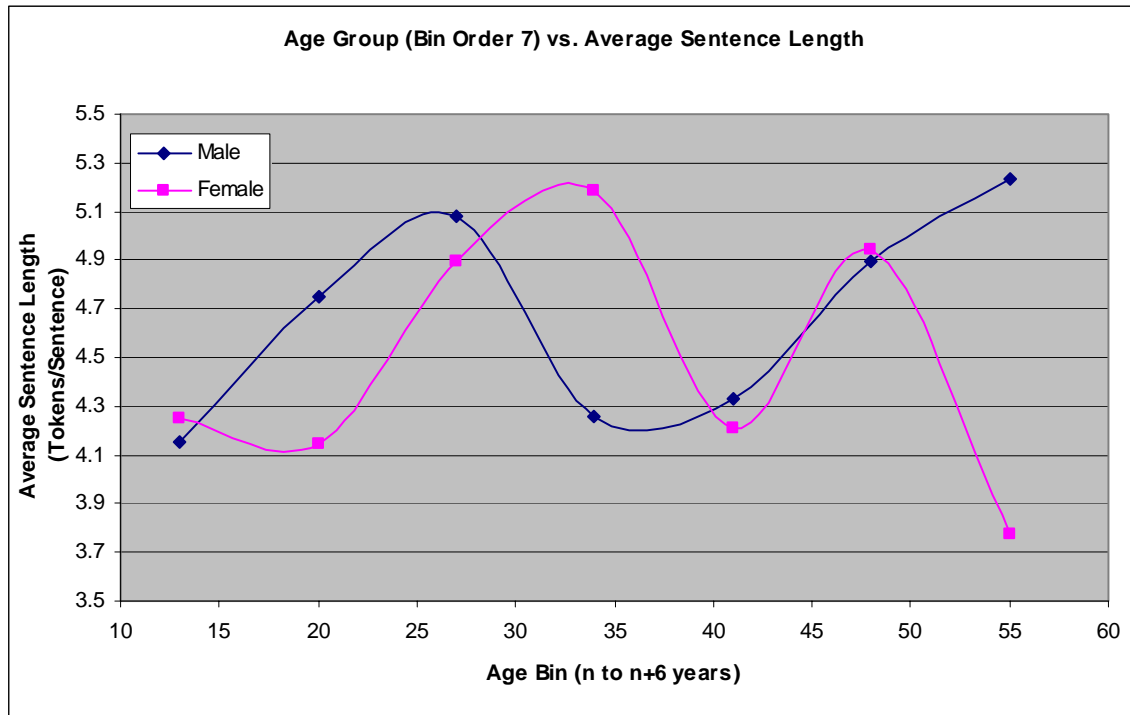


Figure B-7. Age Bin (Order 7) vs. Average Sentence Length

Figure B-8 groups the ages into bins of size 8. Thus, ages 13-20 form the first bin labeled as the point 13, ages 21-28 form the second bin labeled as the point 21, etc.

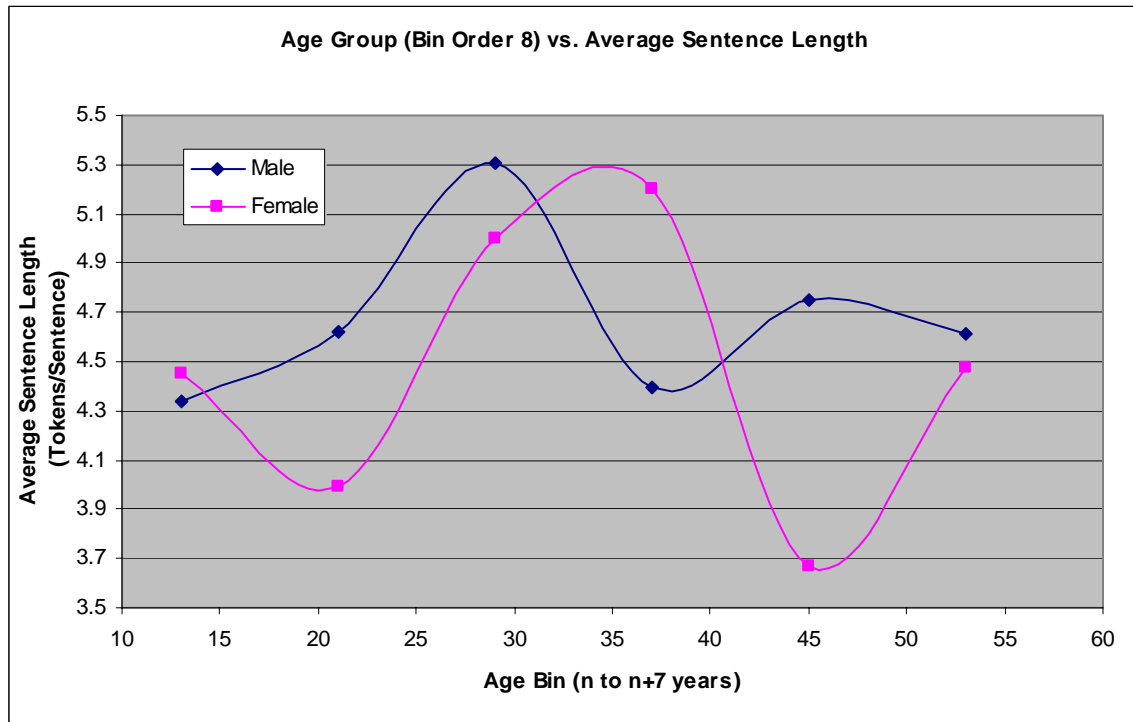


Figure B-8. Age Bin (Order 8) vs. Average Sentence Length

B. MEASURES OF INDIVIDUAL DOCUMENT LENGTHS

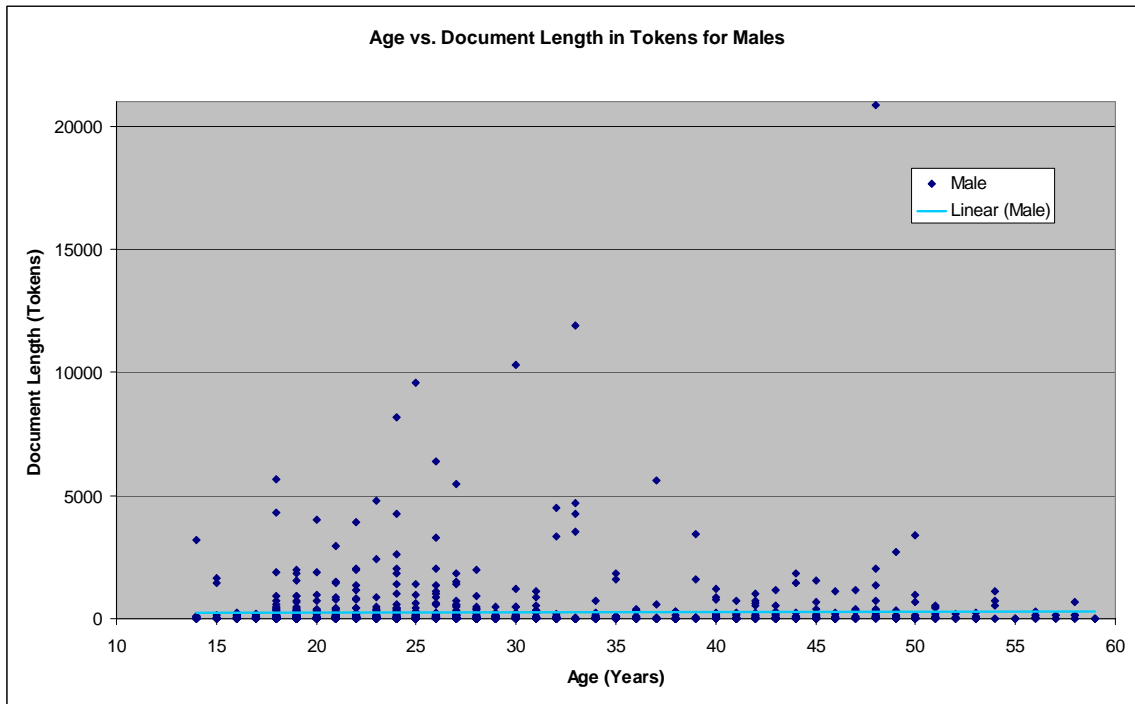


Figure B-9. Age vs. Document Length in Tokens for Males

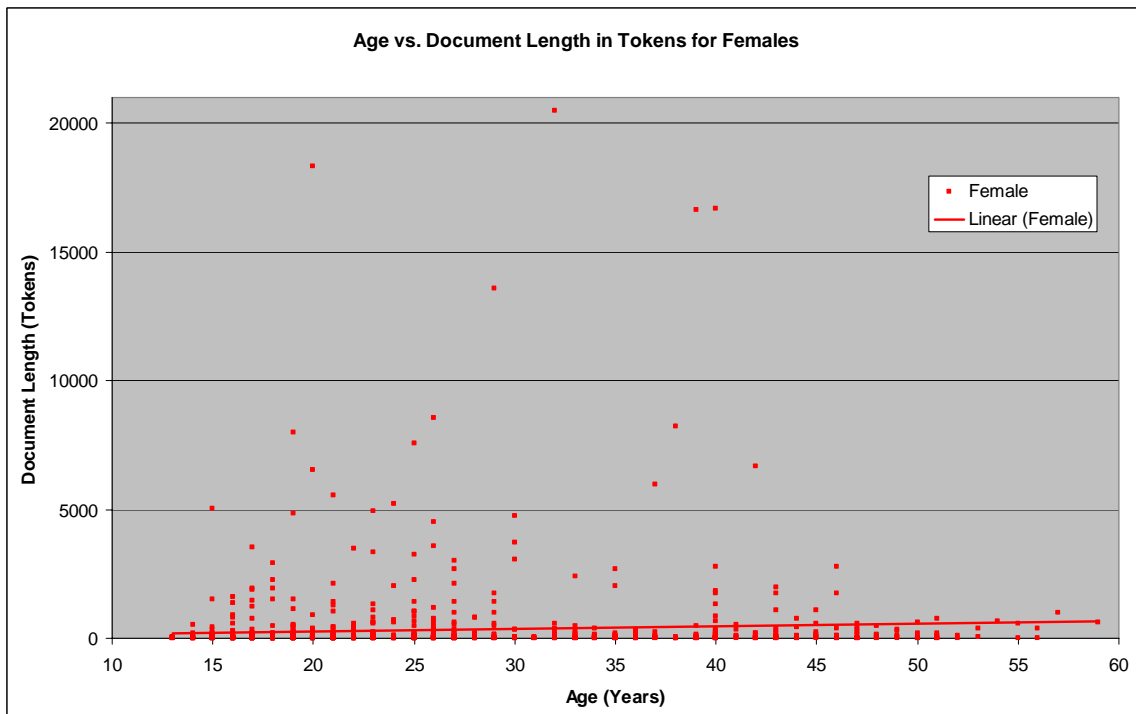


Figure B-10. Age vs. Document Length in Tokens for Females

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C: TABLES FOR VOCABULARY VARIETY

This appendix contains the table of raw counts for types or unique tokens and the table showing the average vocabulary variety.

A. TYPE COUNTS

Types are counts of unique tokens. Note that the sum of types down a column or across a row does not equal to the value shown in the "Total" box. In fact, it is always less than the sum. This makes sense because when separated, say by gender, some types are counted in both the female group and the male group. Thus, when we combine them together in the "Total" box, the duplicate will get discarded, lowering the "Total" type counts.

Age Group	Male	Female	Total
Unspecified	14428	11390	20641
13-19	7361	8782	13071
20-29	14992	16000	24427
30-39	10522	10262	16618
40-49	9051	8004	13798
50-59	3192	1837	4190
Total	34652	32713	67365

Table C-1. Type Count for Training Set

B. AVERAGE VOCABULARY VARIETY

The average vocabulary variety is calculated by dividing the number of types by the number of tokens.

Age Group	Male	Female	Total
Unspecified	0.12132	0.14024	0.26156
13-19	0.17934	0.14064	0.31998
20-29	0.11407	0.10716	0.22123
30-39	0.14873	0.13198	0.28071
40-49	0.16751	0.14873	0.31624
50-59	0.25840	0.31992	0.57832
Total	0.08086	0.07603	0.15689

Table C-2. File Size for Testing Set

APPENDIX D: TABLES AND FIGURES FOR EMOTICONS

This appendix contains the tables and figures of average emoticon tokens and types separated by gender. There are also figures with the genders combined for the average emoticon tokens and types.

A. TABLES

	13-19	20-29	30-39	40-49	50-59	Total
:- (0.00233	0.00128	0.00007	0.00052	0.00039	0.00101
:-)	0.00354	0.00234	0.00079	0.00383	0.01321	0.00288
:-@	0.00040	0.00103	0.00007	0.00009	0.00000	0.00053
:-o	0.00071	0.00131	0.00029	0.00052	0.00078	0.00085
:beer:	0.00223	0.00312	0.00100	0.00183	0.00155	0.00225
:blush:	0.00091	0.00078	0.00021	0.00026	0.00000	0.00056
:love:	0.00304	0.00046	0.00007	0.00009	0.02099	0.00150
:tongue:	0.00111	0.00195	0.00007	0.00104	0.00700	0.00147
;-)	0.00496	0.00238	0.00043	0.00209	0.00544	0.00242
>:->	0.00051	0.00078	0.00000	0.00017	0.00000	0.00044
Total	0.01973	0.01545	0.00301	0.01044	0.04936	0.01391

Table D-1. Emoticon Token per Male Sentence

	13-19	20-29	30-39	40-49	50-59	Total
:- (0.00116	0.00097	0.00138	0.00078	0.00000	0.00103
:-)	0.00143	0.00222	0.00290	0.01327	0.01259	0.00419
:- @	0.00102	0.00097	0.00022	0.00016	0.00000	0.00070
:- o	0.00075	0.00265	0.00109	0.00101	0.00000	0.00170
:beer:	0.00041	0.00066	0.00196	0.00312	0.00157	0.00126
:blush:	0.00102	0.00094	0.00065	0.00101	0.00000	0.00090
:love:	0.00967	0.00174	0.00043	0.00211	0.00079	0.00305
:tongue:	0.00129	0.00094	0.00094	0.00125	0.00079	0.00106
;-)	0.00279	0.00171	0.00101	0.00390	0.00787	0.00225
>:->	0.00014	0.00048	0.00051	0.00133	0.00000	0.00055
Total	0.01968	0.01328	0.01109	0.02795	0.02360	0.01669

Table D-2. Emoticon Token per Female Sentence

	13-19	20-29	30-39	40-49	50-59	Total
:- (0.00233	0.00121	0.00007	0.00052	0.00039	0.00098
:-)	0.00334	0.00217	0.00072	0.00383	0.00700	0.00251
:- @	0.00020	0.00046	0.00007	0.00009	0.00000	0.00026
:- o	0.00071	0.00128	0.00014	0.00052	0.00039	0.00079
:beer:	0.00182	0.00167	0.00050	0.00157	0.00117	0.00141
:blush:	0.00081	0.00078	0.00021	0.00026	0.00000	0.00054
:love:	0.00182	0.00032	0.00007	0.00009	0.01632	0.00107
:tongue:	0.00071	0.00195	0.00007	0.00078	0.00117	0.00113
;-)	0.00486	0.00220	0.00043	0.00209	0.00505	0.00232
>:->	0.00040	0.00039	0.00000	0.00017	0.00000	0.00026
Total	0.01700	0.01243	0.00229	0.00992	0.03148	0.01127

Table D-3. Emoticon Type per Male Sentence

	13-19	20-29	30-39	40-49	50-59	Total
:- (0.00116	0.00080	0.00116	0.00078	0.00000	0.00080
:-)	0.00136	0.00162	0.00283	0.01296	0.01259	0.00162
:- @	0.00089	0.00048	0.00022	0.00016	0.00000	0.00048
:- o	0.00048	0.00239	0.00109	0.00101	0.00000	0.00239
:beer:	0.00041	0.00051	0.00007	0.00273	0.00157	0.00051
:blush:	0.00102	0.00083	0.00065	0.00101	0.00000	0.00083
:love:	0.00552	0.00117	0.00036	0.00086	0.00079	0.00117
:tongue:	0.00109	0.00083	0.00094	0.00117	0.00079	0.00083
; -)	0.00245	0.00160	0.00101	0.00375	0.00787	0.00160
> :->	0.00007	0.00048	0.00036	0.00133	0.00000	0.00048
Total	0.01444	0.01072	0.00870	0.02576	0.02360	0.01072

Table D-4. Emoticon Type per Female Sentence

B. FIGURES

Figure 3-9 shows a summary of the average emoticon token per sentence. To calculate the average emoticon token per sentence, we counted up the number of each emoticon and divided it by the total number of sentences. The age groups is grouped as follows: ages 13-19 is labeled as 15, ages 20-29 is labeled as 25, ages 30-39 is labeled as 35, ages 40-49 is labeled as 45, ages 50-59 is labeled as 50.

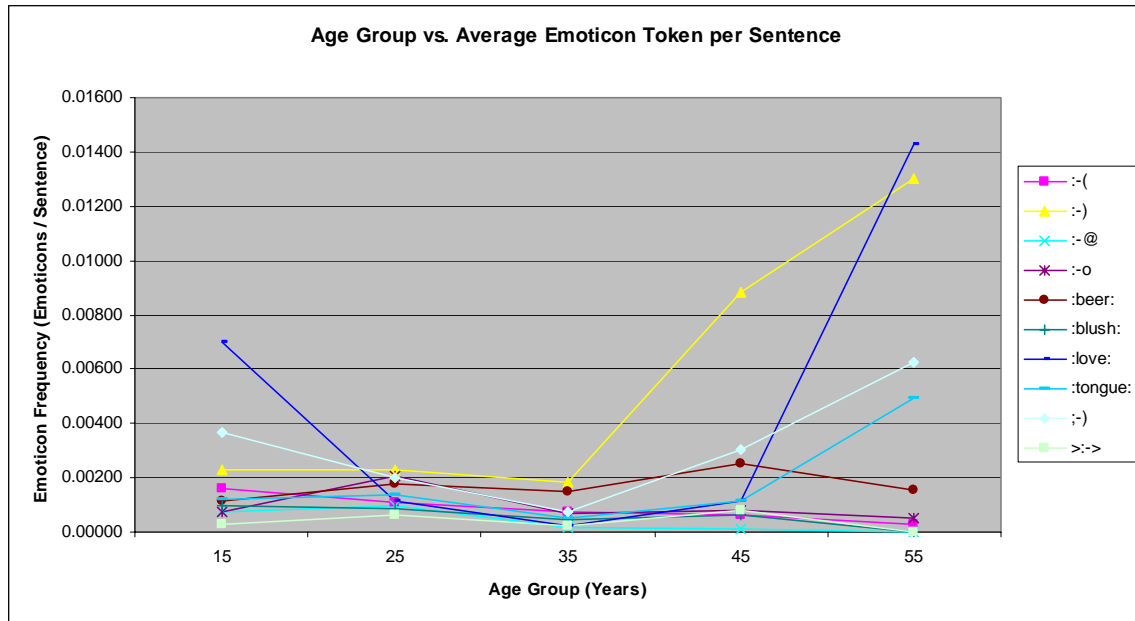


Figure D-1. Age Group vs. Average Emoticon Token per Sentence

To calculate the average emoticon types per sentence, we counted up the number of sentences with a specific emoticon and divided it by the total number of sentences.

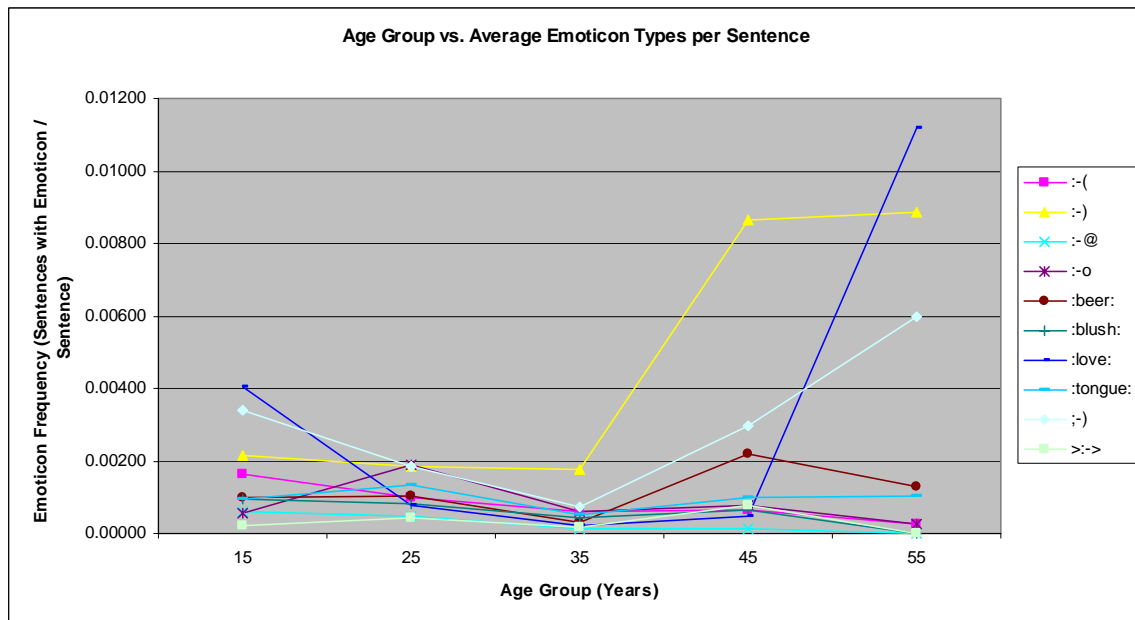


Figure D-2. Age Group vs. Emoticon Types per Sentence

The following two graphs are emoticon types separated by gender.

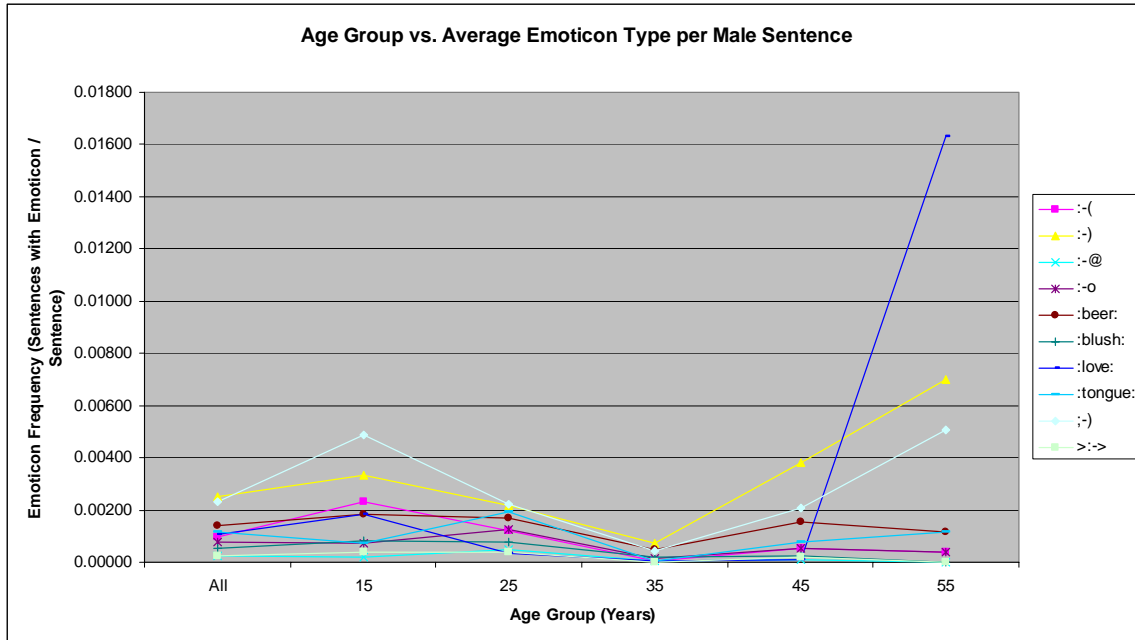


Figure D-3. Age Group vs. Average Emoticon Type per Male Sentence

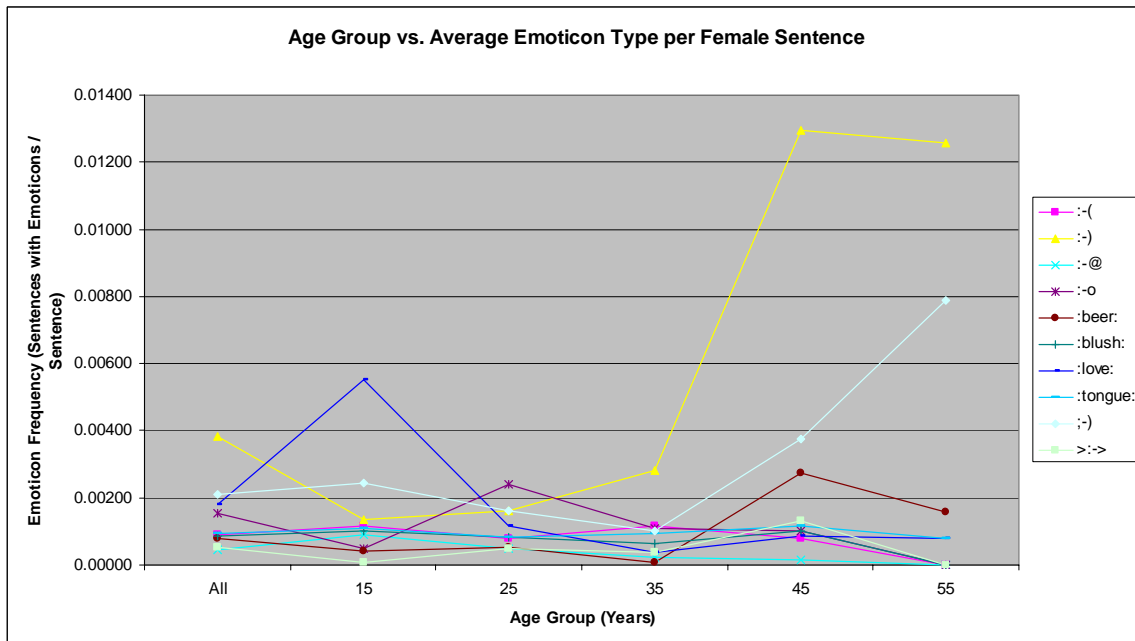


Figure D-4. Age Group vs. Average Emoticon Type per Female Sentence

APPENDIX E: TABLES AND FIGURES FOR PUNCTUATIONS

This appendix contains the tables and figures of average punctuation tokens and types separated by gender. There are also figures with the genders combined for the average punctuation tokens and types.

A. TABLES

Punc.	13-19	20-29	30-39	40-49	50-59	Total
!	0.09532	0.12744	0.07696	0.08264	0.12359	0.07689
"	0.01396	0.01324	0.01382	0.01522	0.02099	0.01044
#	0.00334	0.00611	0.00515	0.00783	0.02332	0.00478
\$	0.00152	0.00117	0.00093	0.00148	0.00039	0.00088
%	0.00040	0.00071	0.00057	0.00026	0.00000	0.00039
&	0.00263	0.00359	0.00100	0.00165	0.00117	0.00182
'	0.06203	0.08156	0.11498	0.08760	0.07890	0.06405
(0.03188	0.02908	0.03608	0.05567	0.85037	0.04995
)	0.04200	0.04325	0.04890	0.08716	0.90711	0.06322
*	0.01477	0.01374	0.00408	0.01253	0.01749	0.00871
+	0.00233	0.00053	0.00072	0.00035	0.03809	0.00168
,	0.04918	0.07947	0.08140	0.08133	0.06918	0.05564
-	0.04179	0.03068	0.02520	0.03149	0.10066	0.02517
.	0.71190	0.54062	0.47544	0.86682	0.94792	0.46200
/	0.01852	0.02301	0.01332	0.00939	0.03653	0.01363
:	0.03198	0.04243	0.03408	0.04332	0.09366	0.03049

;	0.00941	0.00788	0.00494	0.00565	0.00972	0.00530
<	0.00870	0.00934	0.00473	0.01261	0.05558	0.00786
=	0.00273	0.00749	0.00215	0.00374	0.00039	0.00349
>	0.00577	0.00440	0.00115	0.00435	0.00622	0.00294
?	0.11374	0.11356	0.12393	0.13892	0.17256	0.09053
@	0.01730	0.00458	0.00243	0.00548	0.01555	0.00489
[0.00051	0.00057	0.00072	0.00191	0.04780	0.00197
\	0.00111	0.00124	0.00029	0.00113	0.00078	0.00073
]	0.00142	0.00114	0.00079	0.00278	0.06685	0.00292
^	0.00101	0.00199	0.00079	0.00130	0.00233	0.00110
_	0.00557	0.00447	0.00630	0.02288	0.01166	0.00629
`	0.00000	0.00000	0.00007	0.00017	0.00000	0.00003
{	0.00000	0.00000	0.00136	0.00235	0.00000	0.00051
	0.00040	0.00107	0.00029	0.00000	0.00000	0.00043
}	0.00000	0.00007	0.00150	0.00244	0.00000	0.00057
~	0.00000	0.00163	0.00186	0.00235	0.01166	0.00144
Total	1.29124	1.19608	1.08591	1.59281	3.71045	1.00074

Table E-1. Punctuation Tokens per Male Sentence

Punc.	13-19	20-29	30-39	40-49	50-59	Total
!	0.11909	0.18890	0.08747	0.05340	0.04170	0.10917
"	0.01375	0.00986	0.02761	0.00375	0.01023	0.01047
#	0.00620	0.00137	0.00130	0.00827	0.00315	0.00282
\$	0.00014	0.00043	0.00036	0.00070	0.00000	0.00033
%	0.00061	0.00023	0.00123	0.00031	0.00000	0.00040
&	0.00041	0.00080	0.00022	0.00101	0.00000	0.00053
'	0.09642	0.10121	0.11950	0.06932	0.14162	0.08128
(0.01144	0.06846	0.05471	0.14223	0.23603	0.05762
)	0.02431	0.08000	0.06551	0.19454	0.31629	0.07364
*	0.01791	0.00804	0.01087	0.01132	0.00157	0.00891
+	0.00034	0.00123	0.00014	0.00687	0.00000	0.00146
,	0.15273	0.06632	0.11523	0.21772	0.13847	0.09652
-	0.03051	0.05352	0.02384	0.04114	0.04091	0.03421
.	0.53929	0.68484	0.78455	0.90976	0.34146	0.58033
/	0.02261	0.01462	0.00725	0.00422	0.00236	0.01060
:	0.07749	0.03922	0.02696	0.06237	0.04249	0.03955
;	0.00565	0.00550	0.00196	0.01218	0.01180	0.00501
<	0.01655	0.01337	0.00710	0.01257	0.00551	0.01035
=	0.00592	0.00162	0.00225	0.00383	0.00000	0.00237
>	0.00633	0.00775	0.00507	0.00656	0.00393	0.00554
?	0.11542	0.09750	0.11907	0.10695	0.10228	0.08736
@	0.00490	0.00544	0.00283	0.00156	0.00000	0.00341

[0.00232	0.00171	0.00167	0.00008	0.00000	0.00125
\	0.00300	0.00043	0.00123	0.00023	0.00000	0.00084
]	0.00701	0.00182	0.00188	0.00000	0.00000	0.00204
^	0.00204	0.00103	0.00101	0.00101	0.00079	0.00099
_	0.01144	0.00553	0.01181	0.00203	0.02203	0.00612
`	0.00027	0.00125	0.00007	0.00000	0.00000	0.00052
{	0.00000	0.00037	0.01826	0.00039	0.00000	0.00286
	0.00313	0.00034	0.00058	0.00000	0.00000	0.00070
}	0.00000	0.00040	0.02225	0.00039	0.00000	0.00345
~	0.00613	0.00162	0.00283	0.00336	0.00236	0.00245
Total	1.30335	1.46473	1.52663	1.87806	1.46499	1.24311

Table E-2. Average Punctuation Tokens per Female Sentence

Punc.	13-19	20-29	30-39	40-49	50-59	Total
!	0.06092	0.06065	0.04825	0.04219	0.06490	0.04067
"	0.00668	0.00650	0.00666	0.00713	0.01088	0.00506
#	0.00142	0.00178	0.00150	0.00183	0.00544	0.00134
\$	0.00142	0.00107	0.00064	0.00087	0.00039	0.00072
%	0.00040	0.00057	0.00043	0.00009	0.00000	0.00030
&	0.00253	0.00344	0.00093	0.00130	0.00117	0.00171
'	0.05677	0.07116	0.09851	0.07803	0.07501	0.05627
(0.00921	0.00884	0.00802	0.01244	0.04858	0.00805
)	0.01649	0.02081	0.01854	0.04141	0.06607	0.01850
*	0.00931	0.00735	0.00229	0.00592	0.00855	0.00471
+	0.00192	0.00046	0.00021	0.00017	0.00078	0.00044
,	0.03886	0.06516	0.06293	0.06411	0.04897	0.04430
-	0.02429	0.01829	0.01439	0.02070	0.02759	0.01415
.	0.22131	0.20386	0.13395	0.20546	0.21182	0.14212
/	0.01154	0.01314	0.00601	0.00644	0.01866	0.00772
:	0.01730	0.03022	0.01969	0.03801	0.04392	0.02066
;	0.00891	0.00742	0.00494	0.00539	0.00933	0.00506
<	0.00688	0.00447	0.00372	0.00800	0.02332	0.00445
=	0.00263	0.00547	0.00200	0.00209	0.00039	0.00261
>	0.00486	0.00245	0.00086	0.00261	0.00389	0.00189
?	0.08854	0.09946	0.10746	0.10195	0.10144	0.07393
@	0.01660	0.00387	0.00222	0.00531	0.01477	0.00451

[0.00051	0.00057	0.00072	0.00078	0.00389	0.00056
\	0.00061	0.00103	0.00029	0.00096	0.00078	0.00058
]	0.00142	0.00110	0.00079	0.00113	0.00505	0.00092
^	0.00071	0.00128	0.00064	0.00130	0.00233	0.00082
_	0.00435	0.00376	0.00515	0.01888	0.00428	0.00502
`	0.00000	0.00000	0.00007	0.00017	0.00000	0.00003
{	0.00000	0.00000	0.00014	0.00200	0.00000	0.00028
	0.00030	0.00103	0.00029	0.00000	0.00000	0.00040
}	0.00000	0.00007	0.00014	0.00209	0.00000	0.00031
~	0.00000	0.00085	0.00100	0.00096	0.00194	0.00060
Total	0.61668	0.64612	0.55341	0.67971	0.80412	0.46869

Table E-3. Average Punctuation Types per Male Sentence

Punc.	13-19	20-29	30-39	40-49	50-59	Total
!	0.06462	0.10565	0.05160	0.02240	0.02832	0.06020
"	0.00660	0.00473	0.01326	0.00195	0.00551	0.00506
#	0.00143	0.00048	0.00058	0.00304	0.00079	0.00091
\$	0.00007	0.00034	0.00036	0.00016	0.00000	0.00021
%	0.00034	0.00020	0.00116	0.00031	0.00000	0.00034
&	0.00041	0.00074	0.00022	0.00094	0.00000	0.00050
'	0.08614	0.08943	0.09160	0.06417	0.11802	0.07023
(0.00946	0.01288	0.01022	0.01257	0.02282	0.00975
)	0.02090	0.01781	0.01797	0.06542	0.06766	0.02226
*	0.01246	0.00596	0.00529	0.00304	0.00079	0.00534
+	0.00034	0.00123	0.00014	0.00156	0.00000	0.00074
,	0.12502	0.05509	0.08950	0.09235	0.10936	0.06692
-	0.01873	0.03431	0.01616	0.02685	0.03462	0.02211
.	0.27148	0.27859	0.20755	0.18454	0.11015	0.20235
/	0.01205	0.01049	0.00551	0.00328	0.00236	0.00704
:	0.05679	0.02836	0.01978	0.05137	0.03934	0.02972
;	0.00497	0.00504	0.00188	0.01202	0.01180	0.00471
<	0.01491	0.00938	0.00471	0.01023	0.00236	0.00790
=	0.00552	0.00162	0.00174	0.00343	0.00000	0.00218
>	0.00511	0.00507	0.00399	0.00429	0.00236	0.00387
?	0.09717	0.07769	0.09899	0.07728	0.08183	0.06995
@	0.00422	0.00485	0.00283	0.00141	0.00000	0.00306

[0.00211	0.00131	0.00152	0.00008	0.00000	0.00105
\	0.00286	0.00043	0.00123	0.00023	0.00000	0.00081
]	0.00681	0.00145	0.00174	0.00000	0.00000	0.00185
^	0.00191	0.00100	0.00072	0.00094	0.00079	0.00091
_	0.00858	0.00408	0.01015	0.00117	0.01652	0.00471
`	0.00027	0.00017	0.00007	0.00000	0.00000	0.00012
{	0.00000	0.00006	0.00123	0.00023	0.00000	0.00023
	0.00313	0.00034	0.00058	0.00000	0.00000	0.00070
}	0.00000	0.00009	0.00130	0.00023	0.00000	0.00025
~	0.00150	0.00086	0.00130	0.00094	0.00079	0.00088
Total	0.84591	0.75974	0.66490	0.64645	0.65618	0.60685

Table E-4. Average Punctuation Types per Female Sentence

B. FIGURES

Figure 3-9 shows a summary of the average punctuation token per sentence. To calculate the average punctuation token per sentence, we counted up the number of each punctuation and divided it by the total number of sentences. The age groups is grouped as follows: ages 13-19 is labeled as 15, ages 20-29 is labeled as 25, ages 30-39 is labeled as 35, ages 40-49 is labeled as 45, ages 50-59 is labeled as 50.

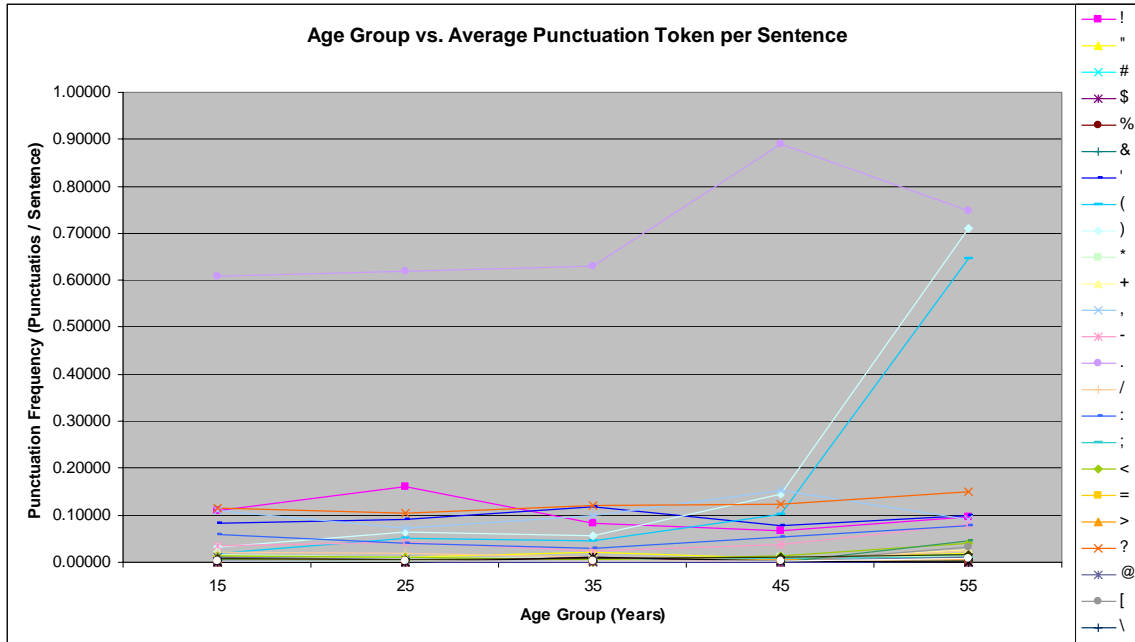


Figure E-1. Age Group vs. Average Punctuation Token per Sentence

To calculate the average emoticon types per sentence, we counted up the number of sentences with a specific emoticon and divided it by the total number of sentences.

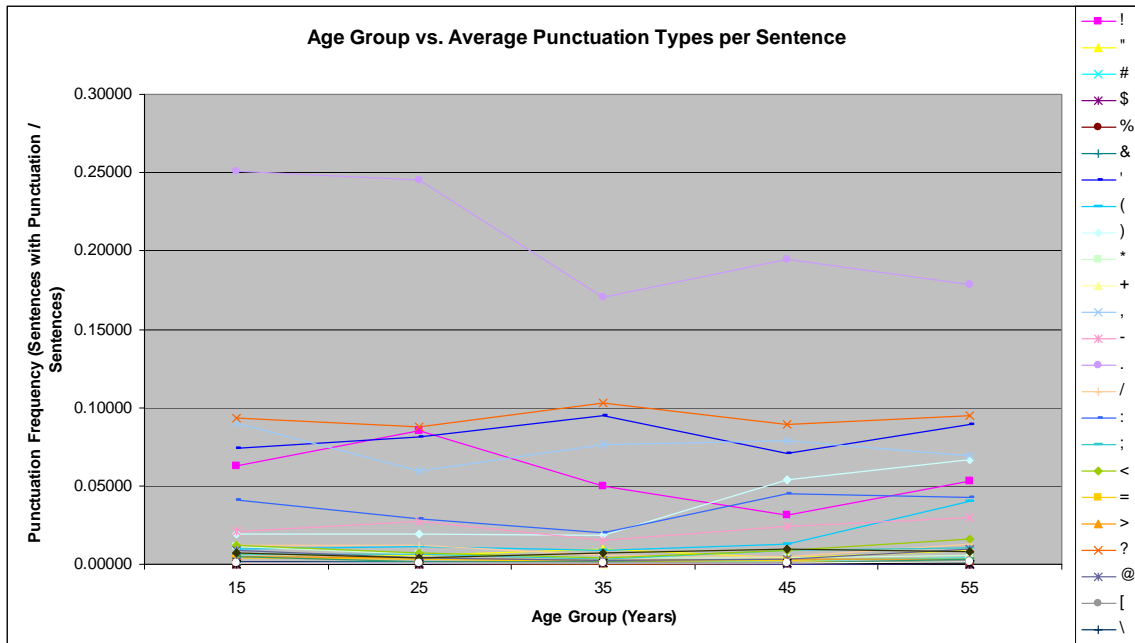


Figure E-2. Age Group vs. Average Punctuation Types per Sentence

The following two graphs are punctuation types separated by gender.

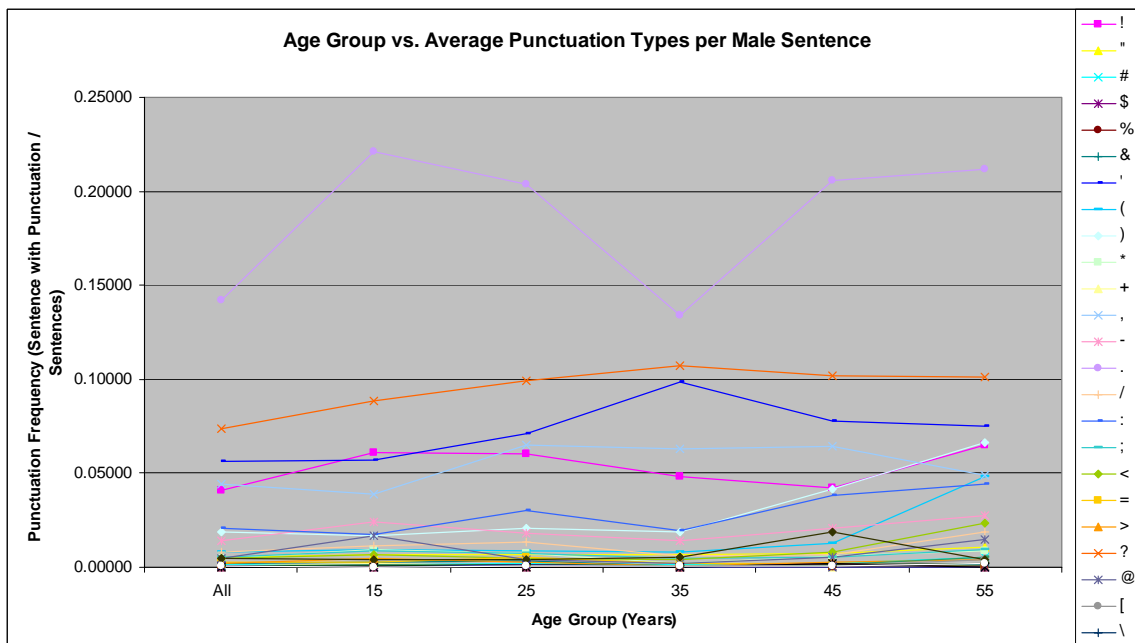


Figure E-3. Age Group vs. Average Punctuation Type per Male Sentence

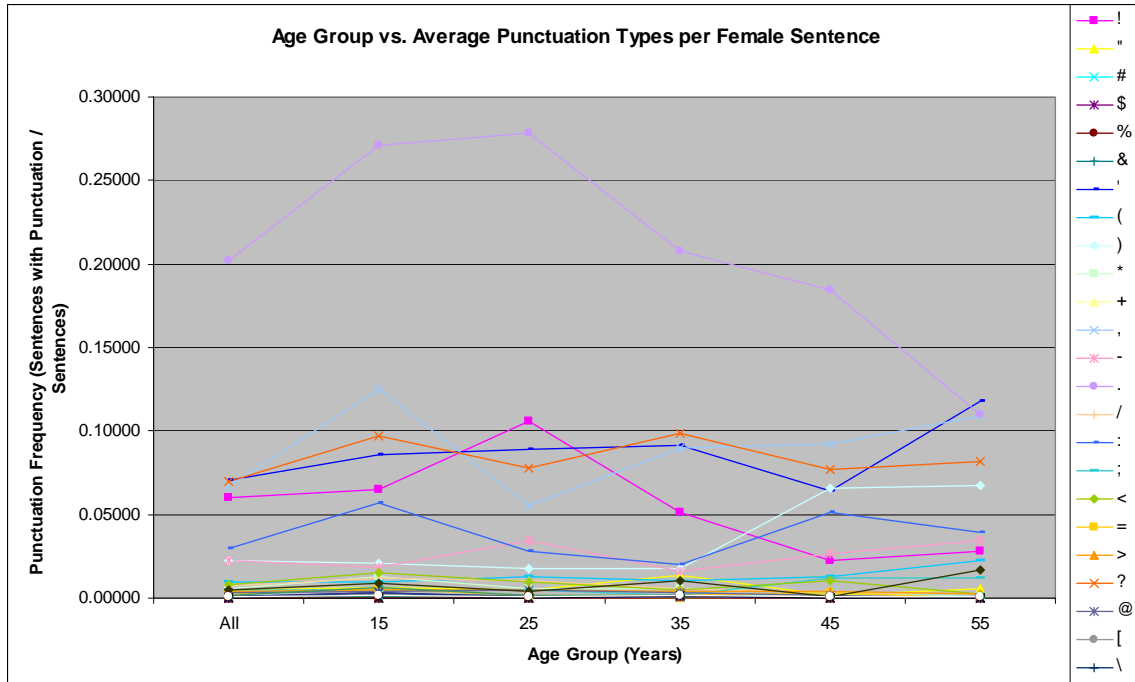


Figure E-4. Age Group vs. Average Punctuation Type per Female Sentence

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX F: PRECISION, RECALL, AND F-SCORES FOR THE FEATURE VECTORS

This appendix contains the precision, recall, and f-scores grouped by the binary gender classification, binary age classification, and multi-class age classification all with and without the prior for six feature vectors: emoticon token, emoticon type, punctuation token, punctuation type, word token, and word type. The key for the feature vectors is included in Appendix H. Feature vectors for which the F-Score do not exist are excluded from the tables.

A. GENDER: BINARY CLASSIFICATION WITH PRIOR

1. All Test Data

Male	Precision	Recall	F-Score
Baseline	0.525835866	1	0.689243028
1	0.527439024	1	0.690618762
2	0.531055901	0.988439306	0.690909091
3	0.526645768	0.971098266	0.682926829
4	0.531365314	0.832369942	0.648648649
5	0.525835866	1	0.689243028
6	0.525835866	1	0.689243028

Table F-1. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.474164134	1	0.643298969
1	1	0.006410256	0.012738854
2	0.714285714	0.032051282	0.061349693
3	0.5	0.032051282	0.060240964
4	0.5	0.185897436	0.271028037

Table F-2. P, R, F-Score for Females

2. Extracted Test Data: Teens and 20s

Male	Precision	Recall	F-Score
Baseline	0.448484848	1	0.619246862
2	0.333333333	0.027027027	0.05
3	0.583333333	0.094594595	0.162790698
4	0.448275862	0.175675676	0.252427184
5	0.355555556	0.216216216	0.268907563
6	0.534883721	0.310810811	0.393162393

Table F-3. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.551515152	1	0.7109375
1	0.543209877	0.967032967	0.695652174
2	0.547169811	0.956043956	0.696
3	0.562091503	0.945054945	0.704918033
4	0.551470588	0.824175824	0.660792952
5	0.516666667	0.681318681	0.587677725
6	0.581967213	0.78021978	0.666666667

Table F-4. P, R, F-Score for Females

3. Extracted Test Data: Teens and 30s

Male	Precision	Recall	F-Score
Baseline	0.428571429	1	0.6
1	0.25	0.022222222	0.040816327
2	0.5	0.022222222	0.042553191
3	0.444444444	0.088888889	0.148148148
4	0.368421053	0.155555556	0.21875

Table F-5. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.571428571	1	0.727272727
1	0.564356436	0.95	0.708074534
2	0.572815534	0.983333333	0.72392638
3	0.572916667	0.916666667	0.705128205
4	0.558139535	0.8	0.657534247
5	0.571428571	1	0.727272727
6	0.571428571	1	0.727272727

Table F-6. P, R, F-Score for Females

4. Extracted Test Data: Teens and 40s

Male	Precision	Recall	F-Score
Baseline	0.42	1	0.591549296
3	0.5	0.023809524	0.045454545
4	0.444444444	0.095238095	0.156862745

Table F-7. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.58	1	0.734177215
1	0.567010309	0.948275862	0.709677419
2	0.567010309	0.948275862	0.709677419
3	0.581632653	0.982758621	0.730769231
4	0.582417582	0.913793103	0.711409396
5	0.58	1	0.734177215
6	0.58	1	0.734177215

Table F-8. P, R, F-Score for Females

5. Extracted Test Data: Teens and 50s

Male	Precision	Recall	F-Score
Baseline	0.384615385	1	0.555555556
3	0.166666667	0.033333333	0.055555556
4	0.5	0.133333333	0.210526316

Table F-9. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.615384615	1	0.761904762
1	0.6	0.9375	0.731707317
2	0.6	0.9375	0.731707317
3	0.597222222	0.895833333	0.716666667
4	0.628571429	0.916666667	0.745762712
5	0.615384615	1	0.761904762
6	0.615384615	1	0.761904762

Table F-10. P, R, F-Score for Females

6. Extracted Test Data: 20s and 30s

Male	Precision	Recall	F-Score
Baseline	0.559701493	1	0.717703349
1	0.553030303	0.973333333	0.70531401
2	0.553846154	0.96	0.702439024
3	0.570247934	0.92	0.704081633
4	0.540322581	0.893333333	0.673366834
5	0.560606061	0.986666667	0.714975845
6	0.559701493	1	0.717703349

Table F-11. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.440298507	1	0.611398964
2	0.25	0.016949153	0.031746032
3	0.538461538	0.118644068	0.194444444
4	0.2	0.033898305	0.057971014
5	0.5	0.016949153	0.032786885

Table F-12. P, R, F-Score for Females

7. Extracted Test Data: 20s and 40s

Male	Precision	Recall	F-Score
Baseline	0.558139535	1	0.71641791
1	0.551181102	0.972222222	0.703517588
2	0.547619048	0.958333333	0.696969697
3	0.559055118	0.986111111	0.713567839
4	0.543103448	0.875	0.670212766
5	0.558139535	1	0.71641791
6	0.558139535	1	0.71641791

Table F-13. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.441860465	1	0.612903226
3	0.5	0.01754386	0.033898305
4	0.307692308	0.070175439	0.114285714

Table F-14. P, R, F-Score for Females

8. Extracted Test Data: 20s and 50s

Male	Precision	Recall	F-Score
Baseline	0.560747664	1	0.718562874
1	0.556603774	0.983333333	0.710843373
2	0.552380952	0.966666667	0.703030303
3	0.564356436	0.95	0.708074534
4	0.558823529	0.95	0.703703704
5	0.560747664	1	0.718562874
6	0.560747664	1	0.718562874

Table F-15. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.439252336	1	0.61038961
3	0.5	0.063829787	0.113207547
4	0.4	0.042553191	0.076923077

Table F-16. P, R, F-Score for Females

9. Extracted Test Data: 30s and 40s

Male	Precision	Recall	F-Score
Baseline	0.623188406	1	0.767857143
1	0.632352941	1	0.774774775
2	0.617647059	0.976744186	0.756756757
3	0.609375	0.906976744	0.728971963
4	0.606557377	0.860465116	0.711538462
5	0.623188406	1	0.767857143
6	0.623188406	1	0.767857143

Table F-17. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.376811594	1	0.547368421
1	1	0.038461538	0.074074074
3	0.2	0.038461538	0.064516129
4	0.25	0.076923077	0.117647059

Table F-18. P, R, F-Score for Females

10. Extracted Test Data: 30s and 50s

Male	Precision	Recall	F-Score
Baseline	0.659574468	1	0.794871795
1	0.652173913	0.967741935	0.779220779
2	0.659574468	1	0.794871795
3	0.674418605	0.935483871	0.783783784
4	0.636363636	0.903225806	0.746666667
5	0.659574468	1	0.794871795
6	0.659574468	1	0.794871795

Table F-19. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.340425532	1	0.507936508
3	0.5	0.125	0.2

Table F-20. P, R, F-Score for Females

11. Extracted Test Data: 40s and 50s

Male	Precision	Recall	F-Score
Baseline	0.666666667	1	0.8
1	0.682926829	1	0.811594203
2	0.682926829	1	0.811594203
3	0.666666667	1	0.8
4	0.658536585	0.964285714	0.782608696
5	0.666666667	1	0.8
6	0.666666667	1	0.8

Table F-21. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.333333333	1	0.5
1	1	0.071428571	0.133333333
2	1	0.071428571	0.133333333

Table F-22. P, R, F-Score for Females

12. Extracted Test Data: Under 26 and 26 or Over

Male	Precision	Recall	F-Score
Baseline	0.512295082	1	0.677506775
1	0.514403292	1	0.679347826
2	0.514644351	0.984	0.675824176
3	0.512820513	0.96	0.668523677
4	0.53960396	0.872	0.666666667
5	0.406779661	0.192	0.260869565
6	0.508403361	0.968	0.666666667

Table F-23. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.487704918	1	0.655647383
1	1	0.008403361	0.016666667
2	0.6	0.025210084	0.048387097
3	0.5	0.042016807	0.07751938
4	0.619047619	0.218487395	0.322981366
5	0.454054054	0.705882353	0.552631579
6	0.333333333	0.016806723	0.032

Table F-24. P, R, F-Score for Females

B. AGE: MULTI-CLASS (5-WAY) CLASSIFICATION WITH PRIOR

1. All Test Data

13-19	Precision	Recall	F-Score
Baseline	0.278688525	1	0.435897436
1	0.5	0.014705882	0.028571429
2	0.25	0.014705882	0.027777778
3	0.2	0.014705882	0.02739726
4	0.266666667	0.058823529	0.096385542
5	0.571428571	0.058823529	0.106666667

Table F-25. P, R, F-Score for Teens

20-29	Precision	Recall	F-Score
Baseline	0.397540984	1	0.568914956
1	0.40167364	0.989690722	0.571428571
2	0.396624473	0.969072165	0.562874251
3	0.397435897	0.958762887	0.561933535
4	0.399122807	0.93814433	0.56
5	0.400843882	0.979381443	0.568862275
6	0.397540984	1	0.568914956

Table F-26. P, R, F-Score for 20s

C. AGE: BINARY CLASSIFICATION WITH PRIOR

1. Extracted Test Data: Teens and 20s

13-19	Precision	Recall	F-Score
Baseline	0.412121212	1	0.583690987
1	0.666666667	0.029411765	0.056338028
2	0.666666667	0.029411765	0.056338028
3	0.833333333	0.073529412	0.135135135
4	0.4	0.058823529	0.102564103
5	0.857142857	0.088235294	0.16

Table F-27. P, R, F-Score for Teens

20-29	Precision	Recall	F-Score
Baseline	0.587878788	1	0.740458015
1	0.592592593	0.989690722	0.741312741
2	0.592592593	0.989690722	0.741312741
3	0.603773585	0.989690722	0.75
4	0.587096774	0.93814433	0.722222222
5	0.607594937	0.989690722	0.752941176
6	0.587878788	1	0.740458015

Table F-28. P, R, F-Score for 20s

2. Extracted Test Data: Teens and 30s

13-19	Precision	Recall	F-Score
Baseline	0.647619048	1	0.786127168
1	0.64	0.941176471	0.761904762
2	0.64	0.941176471	0.761904762
3	0.643564356	0.955882353	0.769230769
4	0.653465347	0.970588235	0.781065089
5	0.647619048	1	0.786127168
6	0.647619048	1	0.786127168

Table F-29. P, R, F-Score for Teens

30-39	Precision	Recall	F-Score
Baseline	0.352380952	1	0.521126761
1	0.2	0.027027027	0.047619048
2	0.2	0.027027027	0.047619048
3	0.25	0.027027027	0.048780488
4	0.5	0.054054054	0.097560976

Table F-30. P, R, F-Score for 20s

3. Extracted Test Data: Teens and 40s

13-19	Precision	Recall	F-Score
Baseline	0.68	1	0.80952381
1	0.666666667	0.941176471	0.780487805
2	0.670103093	0.955882353	0.787878788
3	0.670103093	0.955882353	0.787878788
4	0.686868687	1	0.814371257
5	0.68	1	0.80952381
6	0.68	1	0.80952381

Table F-31. P, R, F-Score for Teens

40-49	Precision	Recall	F-Score
Baseline	0.32	1	0.484848485
4	1	0.03125	0.060606061

Table F-32. P, R, F-Score for 40s

4. Extracted Test Data: Teens and 50s

Teens	Precision	Recall	F-Score
Baseline	0.871794872	1	0.931506849
1	0.864864865	0.941176471	0.901408451
2	0.864864865	0.941176471	0.901408451
3	0.868421053	0.970588235	0.916666667
4	0.871794872	1	0.931506849
5	0.871794872	1	0.931506849
6	0.871794872	1	0.931506849

Table F-33. P, R, F-Score for Teens

5. Extracted Test Data: 20s and 30s

20-29	Precision	Recall	F-Score
Baseline	0.723880597	1	0.83982684
1	0.727272727	0.989690722	0.838427948
2	0.72519084	0.979381443	0.833333333
3	0.732824427	0.989690722	0.842105263
4	0.738461538	0.989690722	0.845814978
5	0.723880597	1	0.83982684
6	0.723880597	1	0.83982684

Table F-34. P, R, F-Score for 20s

30-39	Precision	Recall	F-Score
Baseline	0.276119403	1	0.432748538
1	0.5	0.027027027	0.051282051
2	0.333333333	0.027027027	0.05
3	0.666666667	0.054054054	0.1
4	0.75	0.081081081	0.146341463

Table F-35. P, R, F-Score for 30s

6. Extracted Test Data: 20s and 40s

20-29	Precision	Recall	F-Score
Baseline	0.751937984	1	0.85840708
1	0.748031496	0.979381443	0.848214286
2	0.758064516	0.969072165	0.850678733
3	0.746031746	0.969072165	0.843049327
4	0.753968254	0.979381443	0.852017937
5	0.751937984	1	0.85840708
6	0.751937984	1	0.85840708

Table F-36. P, R, F-Score for 20s

40-49	Precision	Recall	F-Score
Baseline	0.248062016	1	0.397515528
2	0.4	0.0625	0.108108108
4	0.333333333	0.03125	0.057142857

Table F-37. P, R, F-Score for 40s

7. Extracted Test Data: 20s and 50s

20-29	Precision	Recall	F-Score
Baseline	0.906542056	1	0.950980392
1	0.905660377	0.989690722	0.945812808
2	0.904761905	0.979381443	0.940594059
3	0.905660377	0.989690722	0.945812808
4	0.906542056	1	0.950980392
5	0.906542056	1	0.950980392
6	0.906542056	1	0.950980392

Table F-38. P, R, F-Score for 20s

8. Extracted Test Data: 30s and 40s

30-39	Precision	Recall	F-Score
Baseline	0.536231884	1	0.698113208
1	0.536231884	1	0.698113208
2	0.544117647	1	0.704761905
3	0.53125	0.918918919	0.673267327
4	0.538461538	0.945945946	0.68627451
5	0.536231884	1	0.698113208
6	0.710526316	0.72972973	0.72

Table F-39. P, R, F-Score for 30s

40-49	Precision	Recall	F-Score
Baseline	0.463768116	1	0.633663366
2	1	0.03125	0.060606061
3	0.4	0.0625	0.108108108
4	0.5	0.0625	0.111111111
6	0.677419355	0.65625	0.666666667

Table F-40. P, R, F-Score for 40s

9. Extracted Test Data: 30s and 50s

30-39	Precision	Recall	F-Score
Baseline	0.787234043	1	0.880952381
1	0.782608696	0.972972973	0.86746988
2	0.782608696	0.972972973	0.86746988
3	0.772727273	0.918918919	0.839506173
4	0.787234043	1	0.880952381
5	0.787234043	1	0.880952381
6	0.787234043	1	0.880952381

Table F-41. P, R, F-Score for 30s

10. Extracted Test Data: 40s and 50s

40-49	Precision	Recall	F-Score
Baseline	0.761904762	1	0.864864865
1	0.763157895	0.90625	0.828571429
2	0.743589744	0.90625	0.816901408
3	0.761904762	1	0.864864865
4	0.761904762	1	0.864864865
5	0.761904762	1	0.864864865
6	0.761904762	1	0.864864865

Table F-42. P, R, F-Score for 40s

50-59	Precision	Recall	F-Score
Baseline	0.238095238	1	0.384615385
1	0.25	0.1	0.142857143

Table F-43. P, R, F-Score for 50s

11. Extracted Test Data: Under 26 and 26 or Over

< 26	Precision	Recall	F-Score
Baseline	0.540983607	1	0.70212766
1	0.531380753	0.962121212	0.684636119
2	0.533613445	0.962121212	0.686486486
3	0.544303797	0.977272727	0.699186992
4	0.550925926	0.901515152	0.683908046
5	0.53909465	0.992424242	0.698666667
6	0.540983607	1	0.70212766

Table F-44. P, R, F-Score for Under 26

>= 26	Precision	Recall	F-Score
Baseline	0.459016393	1	0.629213483
2	0.166666667	0.008928571	0.016949153
3	0.571428571	0.035714286	0.067226891
4	0.535714286	0.133928571	0.214285714

Table F-45. P, R, F-Score for 26 or older

D. GENDER: BINARY CLASSIFICATION WITHOUT PRIOR

1. All Test Data

Male	Precision	Recall	F-Score
Baseline	0.525835866	1	0.689243028
1	0.530864198	0.994219653	0.692152918
2	0.529595016	0.98265896	0.688259109
3	0.524115756	0.942196532	0.673553719
4	0.5	0.196531792	0.282157676
5	0.444444444	0.069364162	0.12
6	0.556701031	0.312138728	0.4

Table F-46. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.474164134	1	0.643298969
1	0.8	0.025641026	0.049689441
2	0.625	0.032051282	0.06097561
3	0.444444444	0.051282051	0.091954023
4	0.46743295	0.782051282	0.585131894
5	0.466887417	0.903846154	0.615720524
6	0.487068966	0.724358974	0.582474227

Table F-47. P, R, F-Score for Females

2. Extracted Test Data: Teens and 20s

Male	Precision	Recall	F-Score
Baseline	0.448484848	1	0.619246862
1	0.333333333	0.027027027	0.05
2	0.428571429	0.040540541	0.074074074
3	0.6	0.121621622	0.202247191
4	0.472222222	0.22972973	0.309090909
5	0.326923077	0.22972973	0.26984127
6	0.493506494	0.513513514	0.503311258

Table F-48. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.551515152	1	0.7109375
1	0.547169811	0.956043956	0.696
2	0.550632911	0.956043956	0.698795181
3	0.566666667	0.934065934	0.705394191
4	0.558139535	0.791208791	0.654545455
5	0.495575221	0.615384615	0.549019608
6	0.590909091	0.571428571	0.581005587

Table F-49. P, R, F-Score for Females

3. Extracted Test Data: Teens and 30s

Male	Precision	Recall	F-Score
Baseline	0.428571429	1	0.6
1	0.2	0.022222222	0.04
2	0.333333333	0.022222222	0.041666667
3	0.428571429	0.133333333	0.203389831
4	0.4	0.222222222	0.285714286
5	0.266666667	0.177777778	0.213333333
6	0.35	0.311111111	0.329411765

Table F-50. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.571428571	1	0.727272727
1	0.56	0.933333333	0.7
2	0.568627451	0.966666667	0.716049383
3	0.571428571	0.866666667	0.688741722
4	0.5625	0.75	0.642857143
5	0.506666667	0.633333333	0.562962963
6	0.523076923	0.566666667	0.544

Table F-51. P, R, F-Score for Females

4. Extracted Test Data: Teens and 40s

Male	Precision	Recall	F-Score
Baseline	0.42	1	0.591549296
2	0.2	0.023809524	0.042553191
3	0.6	0.071428571	0.127659574
4	0.421052632	0.19047619	0.262295082
5	0.379310345	0.261904762	0.309859155
6	0.375	0.285714286	0.324324324
Male	Precision	Recall	F-Score

Table F-52. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.58	1	0.734177215
1	0.5625	0.931034483	0.701298701
2	0.568421053	0.931034483	0.705882353
3	0.589473684	0.965517241	0.732026144
4	0.580246914	0.810344828	0.676258993
5	0.563380282	0.689655172	0.620155039
6	0.558823529	0.655172414	0.603174603

Table F-53. P, R, F-Score for Females

5. Extracted Test Data: Teens and 50s

Male	Precision	Recall	F-Score
Baseline	0.384615385	1	0.555555556
3	0.375	0.1	0.157894737
4	0.4	0.2	0.266666667
5	0.230769231	0.2	0.214285714
6	0.318181818	0.233333333	0.269230769

Table F-54. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.615384615	1	0.761904762
1	0.594594595	0.916666667	0.721311475
2	0.594594595	0.916666667	0.721311475
3	0.614285714	0.895833333	0.728813559
4	0.619047619	0.8125	0.702702703
5	0.538461538	0.583333333	0.56
6	0.589285714	0.6875	0.634615385

Table F-55. P, R, F-Score for Females

6. Extracted Test Data: 20s and 30s

Male	Precision	Recall	F-Score
Baseline	0.559701493	1	0.717703349
1	0.549618321	0.96	0.699029126
2	0.553846154	0.96	0.702439024
3	0.56302521	0.893333333	0.690721649
4	0.571428571	0.746666667	0.647398844
5	0.515151515	0.226666667	0.314814815
6	0.596491228	0.453333333	0.515151515

Table F-56. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.440298507	1	0.611398964
2	0.25	0.016949153	0.031746032
3	0.466666667	0.118644068	0.189189189
4	0.472222222	0.288135593	0.357894737
5	0.425742574	0.728813559	0.5375
6	0.467532468	0.610169492	0.529411765

Table F-57. P, R, F-Score for Females

7. Extracted Test Data: 20s and 40s

Male	Precision	Recall	F-Score
Baseline	0.558139535	1	0.71641791
1	1	0.013888889	0.02739726
2	0.544	0.944444444	0.69035533
3	0.558333333	0.930555556	0.697916667
4	0.538461538	0.291666667	0.378378378
5	0.545454545	0.25	0.342857143
6	0.675675676	0.347222222	0.458715596

Table F-58. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.441860465	1	0.612903226
1	0.4453125	1	0.616216216
3	0.444444444	0.070175439	0.121212121
4	0.433333333	0.684210526	0.530612245
5	0.4375	0.736842105	0.549019608
6	0.489130435	0.789473684	0.604026846

Table F-59. P, R, F-Score for Females

8. Extracted Test Data: 20s and 50s

Male	Precision	Recall	F-Score
Baseline	0.560747664	1	0.718562874
1	0.542056075	1	0.703030303
2	0.552380952	0.966666667	0.703030303
3	0.568421053	0.9	0.696774194
4	0.568181818	0.833333333	0.675675676
5	0.5	0.233333333	0.318181818
6	0.62745098	0.533333333	0.576576577

Table F-60. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.439252336	1	0.61038961
3	0.5	0.127659574	0.203389831
4	0.473684211	0.191489362	0.272727273
5	0.417721519	0.70212766	0.523809524
6	0.5	0.595744681	0.54368932

Table F-61. P, R, F-Score for Females

9. Extracted Test Data: 30s and 40s

Male	Precision	Recall	F-Score
Baseline	0.623188406	1	0.767857143
1	0.636363636	0.976744186	0.770642202
2	0.625	0.930232558	0.747663551
3	0.603448276	0.813953488	0.693069307
4	0.6	0.837209302	0.699029126
5	0.642857143	0.209302326	0.315789474
6	0.631578947	0.558139535	0.592592593

Table F-62. P, R, F-Score for Males

Feature	Precision	Recall	F-Score
Female	0.376811594	1	0.547368421
1	0.666666667	0.076923077	0.137931034
2	0.4	0.076923077	0.129032258
3	0.272727273	0.115384615	0.162162162
4	0.222222222	0.076923077	0.114285714
5	0.381818182	0.807692308	0.518518519
6	0.387096774	0.461538462	0.421052632

Table F-63. P, R, F-Score for Females

10. Extracted Test Data: 30s and 50s

Male	Precision	Recall	F-Score
Baseline	0.659574468	1	0.794871795
1	0.666666667	0.967741935	0.789473684
2	0.666666667	0.967741935	0.789473684
3	0.72972973	0.870967742	0.794117647
4	0.675	0.870967742	0.76056338
5	0.615384615	0.258064516	0.363636364
6	0.666666667	0.193548387	0.3

Table F-64. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.340425532	1	0.507936508
1	0.5	0.0625	0.111111111
2	0.5	0.0625	0.111111111
3	0.6	0.375	0.461538462
4	0.428571429	0.1875	0.260869565
5	0.323529412	0.6875	0.44
6	0.342105263	0.8125	0.481481481

Table F-65. P, R, F-Score for Females

11. Extracted Test Data: 40s and 50s

Male	Precision	Recall	F-Score
Baseline	0.512295082	1	0.677506775
1	0.514403292	1	0.679347826
2	0.514644351	0.984	0.675824176
3	0.512820513	0.96	0.668523677
4	0.53960396	0.872	0.666666667
5	0.406779661	0.192	0.260869565
6	0.508403361	0.968	0.666666667

Table F-66. P, R, F-Score for Males

Feature	Precision	Recall	F-Score
Female	0.487704918	1	0.655647383
1	1	0.008403361	0.016666667
2	0.6	0.025210084	0.048387097
3	0.5	0.042016807	0.07751938
4	0.619047619	0.218487395	0.322981366
5	0.454054054	0.705882353	0.552631579
6	0.333333333	0.016806723	0.032

Table F-67. P, R, F-Score for Females

12. Extracted Test Data: Under 26 and 26 or Over

Male	Precision	Recall	F-Score
Baseline	0.512295082	1	0.677506775
1	0.514522822	0.992	0.677595628
2	0.512605042	0.976	0.672176309
3	0.515021459	0.96	0.670391061
4	0.526315789	0.16	0.245398773
5	0.406779661	0.192	0.260869565
6	0.53968254	0.272	0.361702128

Table F-68. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.487704918	1	0.655647383
1	0.666666667	0.016806723	0.032786885
2	0.5	0.025210084	0.048
3	0.545454545	0.050420168	0.092307692
4	0.490291262	0.848739496	0.621538462
5	0.454054054	0.705882353	0.552631579
6	0.497237569	0.756302521	0.6

Table F-69. P, R, F-Score for Females

E. AGE: MULTI-CLASS (5-WAY) CLASSIFICATION WITHOUT PRIOR

1. All Test Data

13-19	Precision	Recall	F-Score
Baseline	0.278688525	1	0.435897436
2	1	0.014705882	0.028985507
3	0.5	0.058823529	0.105263158
4	0.296296296	0.117647059	0.168421053
5	0.355932203	0.308823529	0.330708661

Table F-70. P, R, F-Score for Teens

20-29	Precision	Recall	F-Score
Baseline	0.397540984	1	0.568914956
1	0.4	0.969072165	0.56626506
2	0.396551724	0.948453608	0.559270517
3	0.403669725	0.907216495	0.558730159
4	0.38974359	0.783505155	0.520547945
6	0.454545455	0.257731959	0.328947368
20-29	Precision	Recall	F-Score

Table F-71. P, R, F-Score for 20s

30-39	Precision	Recall	F-Score
Baseline	0.151639344	1	0.263345196
4	0.428571429	0.081081081	0.136363636
6	0.206349206	0.702702703	0.319018405

Table F-72. P, R, F-Score for 30s

40-49	Precision	Recall	F-Score
Baseline	0.131147541	1	0.231884058
5	0.152542373	0.84375	0.258373206

Table F-73. P, R, F-Score for 40s

50-59	Precision	Recall	F-Score
Baseline	0.040983607	1	0.078740157
6	0.047619048	0.3	0.082191781

Table F-74. P, R, F-Score for 50s

F. AGE: BINARY CLASSIFICATION WITHOUT PRIOR

1. Extracted Test Data: Teens and 20s

13-19	Precision	Recall	F-Score
Baseline	0.412121212	1	0.583690987
1	0.666666667	0.029411765	0.056338028
2	0.5	0.029411765	0.055555556
3	0.636363636	0.102941176	0.17721519
4	0.419354839	0.191176471	0.262626263
5	0.5	0.382352941	0.433333333
6	0.421686747	0.514705882	0.463576159

Table F-75. P, R, F-Score for Teens

20-29	Precision	Recall	F-Score
Baseline	0.587878788	1	0.740458015
1	0.592592593	0.989690722	0.741312741
2	0.590062112	0.979381443	0.736434109
3	0.603896104	0.958762887	0.741035857
4	0.589552239	0.81443299	0.683982684
5	0.628318584	0.731958763	0.676190476
6	0.597560976	0.505154639	0.547486034

Table F-76. P, R, F-Score for 20s

2. Extracted Test Data: Teens and 30s

13-19	Precision	Recall	F-Score
Baseline	0.647619048	1	0.786127168
1	0.646464646	0.941176471	0.766467066
2	0.646464646	0.941176471	0.766467066
3	0.652631579	0.911764706	0.760736196
4	0.663157895	0.926470588	0.773006135
5	0.722222222	0.382352941	0.5
6	0.75	0.485294118	0.589285714

Table F-77. P, R, F-Score for Teens

30-39	Precision	Recall	F-Score
Baseline	0.352380952	1	0.521126761
1	0.333333333	0.054054054	0.093023256
2	0.333333333	0.054054054	0.093023256
3	0.4	0.108108108	0.170212766
4	0.5	0.135135135	0.212765957
5	0.391304348	0.72972973	0.509433962
6	0.426229508	0.702702703	0.530612245

Table F-78. P, R, F-Score for 30s

3. Extracted Test Data: Teens and 40s

13-19	Precision	Recall	F-Score
Baseline	0.68	1	0.80952381
1	0.663157895	0.926470588	0.773006135
2	0.677419355	0.926470588	0.782608696
3	0.684210526	0.955882353	0.797546012
4	0.681818182	0.882352941	0.769230769
5	0.72972973	0.397058824	0.514285714
6	0.720588235	0.720588235	0.720588235

Table F-79. P, R, F-Score for Teens

40-49	Precision	Recall	F-Score
Baseline	0.32	1	0.484848485
2	0.285714286	0.0625	0.102564103
3	0.4	0.0625	0.108108108
4	0.333333333	0.125	0.181818182
5	0.349206349	0.6875	0.463157895
6	0.40625	0.40625	0.40625

Table F-80. P, R, F-Score for 40s

4. Extracted Test Data: Teens and 50s

13-19	Precision	Recall	F-Score
Baseline	0.871794872	1	0.931506849
1	0.864864865	0.941176471	0.901408451
2	0.863013699	0.926470588	0.893617021
3	0.861111111	0.911764706	0.885714286
4	0.863013699	0.926470588	0.893617021
5	0.807692308	0.308823529	0.446808511
6	0.875	0.720588235	0.790322581

Table F-81. P, R, F-Score for Teens

50-59	Precision	Recall	F-Score
Baseline	0.128205128	1	0.227272727
5	0.096153846	0.5	0.161290323
6	0.136363636	0.3	0.1875

Table F-82. P, R, F-Score for 50s

5. Extracted Test Data: 20s and 30s

20-29	Precision	Recall	F-Score
Baseline	0.723880597	1	0.83982684
1	0.723076923	0.969072165	0.828193833
2	0.730769231	0.979381443	0.837004405
3	0.736	0.948453608	0.828828829
4	0.74789916	0.917525773	0.824074074
5	0.714285714	0.257731959	0.378787879
6	0.816666667	0.505154639	0.624203822

Table F-83. P, R, F-Score for 20s

30-39	Precision	Recall	F-Score
Baseline	0.276119403	1	0.432748538
1	0.25	0.027027027	0.048780488
2	0.5	0.054054054	0.097560976
3	0.444444444	0.108108108	0.173913043
4	0.466666667	0.189189189	0.269230769
5	0.272727273	0.72972973	0.397058824
6	0.351351351	0.702702703	0.468468468

Table F-84. P, R, F-Score for 30s

6. Extracted Test Data: 20s and 40s

20-29	Precision	Recall	F-Score
Baseline	0.751937984	1	0.85840708
1	0.746031746	0.969072165	0.843049327
2	0.762295082	0.958762887	0.849315068
3	0.741666667	0.917525773	0.820276498
4	0.760683761	0.917525773	0.831775701
5	0.805555556	0.298969072	0.436090226
6	0.757575758	0.257731959	0.384615385

Table F-85. P, R, F-Score for 20s

40-49	Precision	Recall	F-Score
Baseline	0.248062016	1	0.397515528
2	0.428571429	0.09375	0.153846154
3	0.111111111	0.03125	0.048780488
4	0.333333333	0.125	0.181818182
5	0.268817204	0.78125	0.4
6	0.25	0.75	0.375

Table F-86. P, R, F-Score for 40s

7. Extracted Test Data: 20s and 50s

20-29	Precision	Recall	F-Score
Baseline	0.751937984	1	0.85840708
1	0.746031746	0.969072165	0.843049327
2	0.762295082	0.958762887	0.849315068
3	0.741666667	0.917525773	0.820276498
4	0.760683761	0.917525773	0.831775701
5	0.805555556	0.298969072	0.436090226
6	0.757575758	0.257731959	0.384615385

Table F-87. P, R, F-Score for 20s

50-59	Precision	Recall	F-Score
Baseline	0.093457944	1	0.170940171
6	0.111111111	0.3	0.162162162

Table F-88. P, R, F-Score for 50s

8. Extracted Test Data: 30s and 40s

30-39	Precision	Recall	F-Score
Baseline	0.536231884	1	0.698113208
1	0.529411765	0.972972973	0.685714286
2	0.546875	0.945945946	0.693069307
3	0.525423729	0.837837838	0.645833333
4	0.516666667	0.837837838	0.639175258
5	0.642857143	0.243243243	0.352941176
6	0.710526316	0.72972973	0.72

Table F-89. P, R, F-Score for 30s

40-49	Precision	Recall	F-Score
Baseline	0.463768116	1	0.633663366
2	0.6	0.09375	0.162162162
3	0.4	0.125	0.19047619
4	0.333333333	0.09375	0.146341463
5	0.490909091	0.84375	0.620689655
6	0.677419355	0.65625	0.666666667

Table F-90. P, R, F-Score for 40s

9. Extracted Test Data: 30s and 50s

30-39	Precision	Recall	F-Score
Baseline	0.787234043	1	0.880952381
1	0.777777778	0.945945946	0.853658537
2	0.772727273	0.918918919	0.839506173
3	0.76744186	0.891891892	0.825
4	0.76744186	0.891891892	0.825
5	0.852941176	0.783783784	0.816901408
6	0.818181818	0.72972973	0.771428571

Table F-91. P, R, F-Score for 30s

50-59	Precision	Recall	F-Score
Baseline	0.212765957	1	0.350877193
5	0.384615385	0.5	0.434782609
6	0.285714286	0.4	0.333333333

Table F-92. P, R, F-Score for 50s

10. Extracted Test Data: 40s and 50s

40-49	Precision	Recall	F-Score
Baseline	0.540983607	1	0.70212766
1	0.531380753	0.962121212	0.684636119
2	0.533613445	0.962121212	0.686486486
3	0.544303797	0.977272727	0.699186992
4	0.550925926	0.901515152	0.683908046
5	0.53909465	0.992424242	0.698666667
6	0.540983607	1	0.70212766

Table F-93. P, R, F-Score for 30s

50-59	Precision	Recall	F-Score
Baseline	0.459016393	1	0.629213483
2	0.166666667	0.008928571	0.016949153
3	0.571428571	0.035714286	0.067226891
4	0.535714286	0.133928571	0.214285714

Table F-94. P, R, F-Score for 40s

11. Extracted Test Data: Under 26 and 26 or Over

>= 26	Precision	Recall	F-Score
Baseline	0.540983607	1	0.70212766
1	0.75	0.022727273	0.044117647
2	0.529661017	0.946969697	0.679347826
3	0.75	0.159090909	0.2625
4	0.742424242	0.371212121	0.494949495
5	0.590909091	0.295454545	0.393939394
6	0.6	0.25	0.352941176

Table F-95. P, R, F-Score for under 26

Feature	Precision	Recall	F-Score
Baseline	0.459016393	1	0.629213483
1	0.4625	0.991071429	0.630681818
2	0.125	0.008928571	0.016666667
3	0.486111111	0.9375	0.640243902
4	0.533707865	0.848214286	0.655172414
5	0.47752809	0.758928571	0.586206897
6	0.476190476	0.803571429	0.598006645

Table F-96. P, R, F-Score for 26 and Older

APPDENDIX G: PRECISION, RECALL, AND F-SCORES FOR THE INDIVIDUAL FEATURES

This appendix contains the precision, recall, and f-scores grouped by the binary gender classification, binary age classification, and multi-class age classification all with and without the prior for 84 individual features: 10 emoticon token, 10 emoticon type, 32 punctuation token, 32 punctuation type, 1 word token, 1 and word type. The key for the features is included in Appendix H. Features for which the F-Score do not exist are excluded from the tables.

A. GENDER: BINARY CLASSIFICATION WITH PRIOR

1. All Test Data

Male	Precision	Recall	F-Score
Baseline	0.525835866	1	0.689243028
1	0.525835866	1	0.689243028
2	0.53	0.919075145	0.67230444
3	0.524390244	0.994219653	0.686626747
4	0.524390244	0.994219653	0.686626747
5	0.525835866	1	0.689243028
6	0.525835866	1	0.689243028
7	0.525835866	1	0.689243028
8	0.525835866	1	0.689243028
9	0.525835866	1	0.689243028
10	0.528301887	0.971098266	0.684317719
11	0.527607362	0.994219653	0.689378758
12	0.525835866	1	0.689243028
13	0.525835866	1	0.689243028
14	0.524390244	0.994219653	0.686626747
15	0.527439024	1	0.690618762
16	0.527607362	0.994219653	0.689378758
17	0.525835866	1	0.689243028
18	0.530864198	0.994219653	0.692152918
19	0.525835866	1	0.689243028
20	0.527439024	1	0.690618762
21	0.525835866	1	0.689243028
22	0.525993884	0.994219653	0.688

23	0.525835866	1	0.689243028
24	0.525835866	1	0.689243028
25	0.527439024	1	0.690618762
26	0.529051988	1	0.692
27	0.525835866	1	0.689243028
28	0.526153846	0.988439306	0.686746988
29	0.525835866	1	0.689243028
30	0.525835866	1	0.689243028
31	0.525835866	1	0.689243028
32	0.525835866	1	0.689243028
33	0.525835866	1	0.689243028
34	0.525835866	1	0.689243028
35	0.52293578	0.988439306	0.684
36	0.525835866	1	0.689243028
37	0.525835866	1	0.689243028
38	0.529051988	1	0.692
39	0.525835866	1	0.689243028
40	0.525835866	1	0.689243028
41	0.525835866	1	0.689243028
42	0.525835866	1	0.689243028
43	0.525835866	1	0.689243028
44	0.525835866	1	0.689243028
45	0.525835866	1	0.689243028
46	0.525835866	1	0.689243028
47	0.527439024	1	0.690618762
48	0.525835866	1	0.689243028
49	0.525835866	1	0.689243028
50	0.525835866	1	0.689243028
51	0.525835866	1	0.689243028
52	0.525835866	1	0.689243028
53	0.525835866	1	0.689243028
54	0.525835866	1	0.689243028
55	0.525835866	1	0.689243028
56	0.529595016	0.98265896	0.688259109
57	0.525835866	1	0.689243028
58	0.525835866	1	0.689243028
59	0.525835866	1	0.689243028
60	0.525835866	1	0.689243028
61	0.525835866	1	0.689243028
62	0.525835866	1	0.689243028
63	0.525835866	1	0.689243028
64	0.525835866	1	0.689243028
65	0.525835866	1	0.689243028
66	0.525835866	1	0.689243028
67	0.525835866	1	0.689243028

68	0.525835866	1	0.689243028
69	0.527439024	1	0.690618762
70	0.525835866	1	0.689243028
71	0.525835866	1	0.689243028
72	0.525835866	1	0.689243028
73	0.52293578	0.988439306	0.684
74	0.524390244	0.994219653	0.686626747
75	0.525835866	1	0.689243028
76	0.525835866	1	0.689243028
77	0.525835866	1	0.689243028
78	0.525835866	1	0.689243028
79	0.525835866	1	0.689243028
80	0.525835866	1	0.689243028
81	0.527439024	1	0.690618762
82	0.525835866	1	0.689243028
83	0.525835866	1	0.689243028
84	0.525835866	1	0.689243028

Table G-1. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.474164134	1	0.643298969
2	0.517241379	0.096153846	0.162162162
10	0.545454545	0.038461538	0.071856287
11	0.666666667	0.012820513	0.025157233
15	1	0.006410256	0.012738854
16	0.666666667	0.012820513	0.025157233
18	0.8	0.025641026	0.049689441
20	1	0.006410256	0.012738854
22	0.5	0.006410256	0.012658228
25	1	0.006410256	0.012738854
26	1	0.012820513	0.025316456
28	0.5	0.012820513	0.025
38	1	0.012820513	0.025316456
47	1	0.006410256	0.012738854
56	0.625	0.032051282	0.06097561
69	1	0.006410256	0.012738854
81	1	0.006410256	0.012738854

Table G-2. P, R, F-Score for Females

2. Extracted Test Data: Teens and 20s

Male	Precision	Recall	F-Score
Baseline	0.448484848	1	0.619246862
10	0.428571429	0.040540541	0.074074074
11	0.75	0.040540541	0.076923077
15	0.285714286	0.027027027	0.049382716
16	0.545454545	0.081081081	0.141176471
18	1	0.013513514	0.026666667
24	0.5	0.013513514	0.026315789
28	1	0.013513514	0.026666667
31	1	0.013513514	0.026666667
32	1	0.013513514	0.026666667
33	0.5	0.013513514	0.026315789
34	0.333333333	0.013513514	0.025974026
42	1	0.013513514	0.026666667
53	1	0.013513514	0.026666667
55	0.5	0.013513514	0.026315789
57	0.5	0.013513514	0.026315789
62	0.5	0.027027027	0.051282051
64	0.416666667	0.067567568	0.11627907
71	1	0.013513514	0.026666667
73	1	0.013513514	0.026666667
77	0.355555556	0.216216216	0.268907563
78	0.534883721	0.310810811	0.393162393

Table G-3. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.551515152	1	0.7109375
1	0.551515152	1	0.7109375
2	0.551515152	1	0.7109375
3	0.551515152	1	0.7109375
4	0.551515152	1	0.7109375
5	0.551515152	1	0.7109375
6	0.551515152	1	0.7109375
7	0.551515152	1	0.7109375
8	0.551515152	1	0.7109375
9	0.551515152	1	0.7109375
10	0.550632911	0.956043956	0.698795181
11	0.559006211	0.989010989	0.714285714
12	0.543209877	0.967032967	0.695652174
13	0.551515152	1	0.7109375
14	0.551515152	1	0.7109375
15	0.544303797	0.945054945	0.690763052
16	0.558441558	0.945054945	0.702040816
17	0.551515152	1	0.7109375
18	0.554878049	1	0.71372549
19	0.551515152	1	0.7109375
20	0.551515152	1	0.7109375
21	0.551515152	1	0.7109375
22	0.551515152	1	0.7109375
23	0.548780488	0.989010989	0.705882353
24	0.552147239	0.989010989	0.708661417
25	0.548780488	0.989010989	0.705882353
26	0.548780488	0.989010989	0.705882353
27	0.551515152	1	0.7109375
28	0.554878049	1	0.71372549
29	0.548780488	0.989010989	0.705882353
30	0.551515152	1	0.7109375
31	0.554878049	1	0.71372549
32	0.554878049	1	0.71372549
33	0.552147239	0.989010989	0.708661417
34	0.549382716	0.978021978	0.703557312
35	0.548780488	0.989010989	0.705882353
36	0.551515152	1	0.7109375
37	0.548780488	0.989010989	0.705882353
38	0.548780488	0.989010989	0.705882353
39	0.551515152	1	0.7109375
40	0.551515152	1	0.7109375
41	0.551515152	1	0.7109375
42	0.554878049	1	0.71372549

43	0.551515152	1	0.7109375
44	0.551515152	1	0.7109375
45	0.551515152	1	0.7109375
46	0.551515152	1	0.7109375
47	0.551515152	1	0.7109375
48	0.548780488	0.989010989	0.705882353
49	0.551515152	1	0.7109375
50	0.551515152	1	0.7109375
51	0.551515152	1	0.7109375
52	0.551515152	1	0.7109375
53	0.554878049	1	0.71372549
54	0.551515152	1	0.7109375
55	0.552147239	0.989010989	0.708661417
56	0.551515152	1	0.7109375
57	0.552147239	0.989010989	0.708661417
58	0.551515152	1	0.7109375
59	0.551515152	1	0.7109375
60	0.551515152	1	0.7109375
61	0.551515152	1	0.7109375
62	0.552795031	0.978021978	0.706349206
63	0.548780488	0.989010989	0.705882353
64	0.549019608	0.923076923	0.68852459
65	0.551515152	1	0.7109375
66	0.551515152	1	0.7109375
67	0.551515152	1	0.7109375
68	0.551515152	1	0.7109375
69	0.548780488	0.989010989	0.705882353
70	0.548780488	0.989010989	0.705882353
71	0.554878049	1	0.71372549
72	0.551515152	1	0.7109375
73	0.554878049	1	0.71372549
74	0.551515152	1	0.7109375
75	0.551515152	1	0.7109375
76	0.551515152	1	0.7109375
77	0.516666667	0.681318681	0.587677725
78	0.581967213	0.78021978	0.666666667
79	0.551515152	1	0.7109375
80	0.551515152	1	0.7109375
81	0.551515152	1	0.7109375
82	0.551515152	1	0.7109375
83	0.551515152	1	0.7109375
84	0.551515152	1	0.7109375

Table G-4. P, R, F-Score for Females

3. Extracted Test Data: Teens and 30s

Male	Precision	Recall	F-Score
Baseline	0.428571429	1	0.6
1	1	0.022222222	0.043478261
3	1	0.022222222	0.043478261
4	1	0.022222222	0.043478261
9	0.5	0.022222222	0.042553191
10	0.5	0.022222222	0.042553191
11	0.5	0.044444444	0.081632653
12	0.25	0.022222222	0.040816327
14	0.5	0.044444444	0.081632653
15	0.5	0.044444444	0.081632653
16	0.8	0.088888889	0.16
35	1	0.022222222	0.043478261
37	0.5	0.022222222	0.042553191
38	0.5	0.022222222	0.042553191
42	1	0.022222222	0.043478261
62	0.25	0.022222222	0.040816327
64	0.333333333	0.022222222	0.041666667

Table G-5. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.571428571	1	0.727272727
1	0.576923077	1	0.731707317
2	0.567307692	0.983333333	0.719512195
3	0.576923077	1	0.731707317
4	0.576923077	1	0.731707317
5	0.571428571	1	0.727272727
6	0.571428571	1	0.727272727
7	0.571428571	1	0.727272727
8	0.571428571	1	0.727272727
9	0.572815534	0.983333333	0.72392638
10	0.572815534	0.983333333	0.72392638
11	0.574257426	0.966666667	0.720496894
12	0.564356436	0.95	0.708074534
13	0.571428571	1	0.727272727
14	0.574257426	0.966666667	0.720496894
15	0.574257426	0.966666667	0.720496894
16	0.59	0.983333333	0.7375
17	0.571428571	1	0.727272727
18	0.571428571	1	0.727272727
19	0.571428571	1	0.727272727
20	0.567307692	0.983333333	0.719512195
21	0.571428571	1	0.727272727
22	0.571428571	1	0.727272727
23	0.571428571	1	0.727272727
24	0.571428571	1	0.727272727
25	0.567307692	0.983333333	0.719512195
26	0.571428571	1	0.727272727
27	0.571428571	1	0.727272727
28	0.571428571	1	0.727272727
29	0.567307692	0.983333333	0.719512195
30	0.567307692	0.983333333	0.719512195
31	0.571428571	1	0.727272727
32	0.571428571	1	0.727272727
33	0.567307692	0.983333333	0.719512195
34	0.571428571	1	0.727272727
35	0.576923077	1	0.731707317
36	0.571428571	1	0.727272727
37	0.572815534	0.983333333	0.72392638
38	0.572815534	0.983333333	0.72392638
39	0.571428571	1	0.727272727
40	0.571428571	1	0.727272727
41	0.571428571	1	0.727272727
42	0.576923077	1	0.731707317

43	0.571428571	1	0.727272727
44	0.571428571	1	0.727272727
45	0.571428571	1	0.727272727
46	0.571428571	1	0.727272727
47	0.571428571	1	0.727272727
48	0.567307692	0.983333333	0.719512195
49	0.571428571	1	0.727272727
50	0.571428571	1	0.727272727
51	0.571428571	1	0.727272727
52	0.571428571	1	0.727272727
53	0.567307692	0.983333333	0.719512195
54	0.571428571	1	0.727272727
55	0.567307692	0.983333333	0.719512195
56	0.571428571	1	0.727272727
57	0.567307692	0.983333333	0.719512195
58	0.571428571	1	0.727272727
59	0.571428571	1	0.727272727
60	0.571428571	1	0.727272727
61	0.571428571	1	0.727272727
62	0.564356436	0.95	0.708074534
63	0.567307692	0.983333333	0.719512195
64	0.568627451	0.966666667	0.716049383
65	0.571428571	1	0.727272727
66	0.571428571	1	0.727272727
67	0.571428571	1	0.727272727
68	0.571428571	1	0.727272727
69	0.567307692	0.983333333	0.719512195
70	0.567307692	0.983333333	0.719512195
71	0.567307692	0.983333333	0.719512195
72	0.571428571	1	0.727272727
73	0.571428571	1	0.727272727
74	0.571428571	1	0.727272727
75	0.571428571	1	0.727272727
76	0.571428571	1	0.727272727
77	0.571428571	1	0.727272727
78	0.571428571	1	0.727272727
79	0.571428571	1	0.727272727
80	0.571428571	1	0.727272727
81	0.571428571	1	0.727272727
82	0.571428571	1	0.727272727
83	0.571428571	1	0.727272727
84	0.571428571	1	0.727272727

Table G-6. P, R, F-Score for Females

4. Extracted Test Data: Teens and 40s

Male	Precision	Recall	F-Score
Baseline	0.42	1	0.591549296
11	1	0.023809524	0.046511628
16	0.75	0.071428571	0.130434783
26	0.5	0.023809524	0.045454545
39	1	0.023809524	0.046511628
42	1	0.023809524	0.046511628
59	1	0.023809524	0.046511628
62	0.333333333	0.023809524	0.044444444
64	0.333333333	0.023809524	0.044444444
68	1	0.023809524	0.046511628
71	1	0.023809524	0.046511628

Table G-7.

P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.58	1	0.734177215
1	0.58	1	0.734177215
2	0.58	1	0.734177215
3	0.58	1	0.734177215
4	0.58	1	0.734177215
5	0.58	1	0.734177215
6	0.58	1	0.734177215
7	0.58	1	0.734177215
8	0.58	1	0.734177215
9	0.575757576	0.982758621	0.72611465
10	0.575757576	0.982758621	0.72611465
11	0.585858586	1	0.738853503
12	0.575757576	0.982758621	0.72611465
13	0.58	1	0.734177215
14	0.575757576	0.982758621	0.72611465
15	0.58	1	0.734177215
16	0.59375	0.982758621	0.74025974
17	0.58	1	0.734177215
18	0.58	1	0.734177215
19	0.58	1	0.734177215
20	0.58	1	0.734177215
21	0.58	1	0.734177215
22	0.58	1	0.734177215
23	0.58	1	0.734177215
24	0.58	1	0.734177215
25	0.575757576	0.982758621	0.72611465
26	0.581632653	0.982758621	0.730769231
27	0.58	1	0.734177215
28	0.58	1	0.734177215
29	0.575757576	0.982758621	0.72611465
30	0.575757576	0.982758621	0.72611465
31	0.58	1	0.734177215
32	0.58	1	0.734177215
33	0.575757576	0.982758621	0.72611465
34	0.58	1	0.734177215
35	0.58	1	0.734177215
36	0.58	1	0.734177215
37	0.575757576	0.982758621	0.72611465
38	0.575757576	0.982758621	0.72611465
39	0.585858586	1	0.738853503
40	0.58	1	0.734177215
41	0.58	1	0.734177215
42	0.585858586	1	0.738853503

43	0.58	1	0.734177215
44	0.58	1	0.734177215
45	0.58	1	0.734177215
46	0.575757576	0.982758621	0.72611465
47	0.58	1	0.734177215
48	0.575757576	0.982758621	0.72611465
49	0.58	1	0.734177215
50	0.58	1	0.734177215
51	0.58	1	0.734177215
52	0.58	1	0.734177215
53	0.58	1	0.734177215
54	0.58	1	0.734177215
55	0.58	1	0.734177215
56	0.58	1	0.734177215
57	0.575757576	0.982758621	0.72611465
58	0.58	1	0.734177215
59	0.585858586	1	0.738853503
60	0.58	1	0.734177215
61	0.58	1	0.734177215
62	0.577319588	0.965517241	0.722580645
63	0.575757576	0.982758621	0.72611465
64	0.577319588	0.965517241	0.722580645
65	0.58	1	0.734177215
66	0.58	1	0.734177215
67	0.58	1	0.734177215
68	0.585858586	1	0.738853503
69	0.575757576	0.982758621	0.72611465
70	0.575757576	0.982758621	0.72611465
71	0.585858586	1	0.738853503
72	0.58	1	0.734177215
73	0.58	1	0.734177215
74	0.58	1	0.734177215
75	0.58	1	0.734177215
76	0.58	1	0.734177215
77	0.58	1	0.734177215
78	0.58	1	0.734177215
79	0.58	1	0.734177215
80	0.58	1	0.734177215
81	0.58	1	0.734177215
82	0.58	1	0.734177215
83	0.58	1	0.734177215
84	0.58	1	0.734177215

Table G-8. P, R, F-Score for Females

5. Extracted Test Data: Teens and 50s

Male	Precision	Recall	F-Score
Baseline	0.384615385	1	0.555555556
11	0.75	0.1	0.176470588
12	0.666666667	0.066666667	0.121212121
16	0.75	0.1	0.176470588
42	1	0.033333333	0.064516129
52	1	0.033333333	0.064516129
62	0.333333333	0.033333333	0.060606061
71	1	0.033333333	0.064516129

Table G-9.

P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.615384615	1	0.761904762
1	0.615384615	1	0.761904762
2	0.615384615	1	0.761904762
3	0.615384615	1	0.761904762
4	0.615384615	1	0.761904762
5	0.615384615	1	0.761904762
6	0.615384615	1	0.761904762
7	0.615384615	1	0.761904762
8	0.615384615	1	0.761904762
9	0.61038961	0.979166667	0.752
10	0.61038961	0.979166667	0.752
11	0.635135135	0.979166667	0.770491803
12	0.626666667	0.979166667	0.764227642
13	0.615384615	1	0.761904762
14	0.61038961	0.979166667	0.752
15	0.615384615	1	0.761904762
16	0.635135135	0.979166667	0.770491803
17	0.615384615	1	0.761904762
18	0.615384615	1	0.761904762
19	0.615384615	1	0.761904762
20	0.615384615	1	0.761904762
21	0.615384615	1	0.761904762
22	0.615384615	1	0.761904762
23	0.61038961	0.979166667	0.752
24	0.61038961	0.979166667	0.752
25	0.61038961	0.979166667	0.752
26	0.61038961	0.979166667	0.752
27	0.615384615	1	0.761904762
28	0.615384615	1	0.761904762
29	0.61038961	0.979166667	0.752
30	0.61038961	0.979166667	0.752
31	0.615384615	1	0.761904762
32	0.615384615	1	0.761904762
33	0.61038961	0.979166667	0.752
34	0.615384615	1	0.761904762
35	0.61038961	0.979166667	0.752
36	0.615384615	1	0.761904762
37	0.61038961	0.979166667	0.752
38	0.61038961	0.979166667	0.752
39	0.615384615	1	0.761904762
40	0.615384615	1	0.761904762
41	0.615384615	1	0.761904762
42	0.623376623	1	0.768

43	0.615384615	1	0.761904762
44	0.615384615	1	0.761904762
45	0.615384615	1	0.761904762
46	0.61038961	0.979166667	0.752
47	0.615384615	1	0.761904762
48	0.61038961	0.979166667	0.752
49	0.615384615	1	0.761904762
50	0.615384615	1	0.761904762
51	0.615384615	1	0.761904762
52	0.623376623	1	0.768
53	0.615384615	1	0.761904762
54	0.615384615	1	0.761904762
55	0.61038961	0.979166667	0.752
56	0.615384615	1	0.761904762
57	0.61038961	0.979166667	0.752
58	0.615384615	1	0.761904762
59	0.615384615	1	0.761904762
60	0.615384615	1	0.761904762
61	0.615384615	1	0.761904762
62	0.613333333	0.958333333	0.74796748
63	0.61038961	0.979166667	0.752
64	0.615384615	1	0.761904762
65	0.615384615	1	0.761904762
66	0.615384615	1	0.761904762
67	0.615384615	1	0.761904762
68	0.615384615	1	0.761904762
69	0.61038961	0.979166667	0.752
70	0.61038961	0.979166667	0.752
71	0.623376623	1	0.768
72	0.615384615	1	0.761904762
73	0.615384615	1	0.761904762
74	0.615384615	1	0.761904762
75	0.615384615	1	0.761904762
76	0.615384615	1	0.761904762
77	0.615384615	1	0.761904762
78	0.615384615	1	0.761904762
79	0.615384615	1	0.761904762
80	0.615384615	1	0.761904762
81	0.615384615	1	0.761904762
82	0.615384615	1	0.761904762
83	0.615384615	1	0.761904762
84	0.615384615	1	0.761904762

Table G-10. P, R, F-Score for Females

6. Extracted Test Data: 20s and 30s

Male	Precision	Recall	F-Score
Baseline	0.559701493	1	0.717703349
1	0.557251908	0.973333333	0.708737864
2	0.556390977	0.986666667	0.711538462
3	0.556390977	0.986666667	0.711538462
4	0.556390977	0.986666667	0.711538462
5	0.559701493	1	0.717703349
6	0.559701493	1	0.717703349
7	0.559701493	1	0.717703349
8	0.559701493	1	0.717703349
9	0.551181102	0.933333333	0.693069307
10	0.559701493	1	0.717703349
11	0.568181818	1	0.724637681
12	0.560606061	0.986666667	0.714975845
13	0.559701493	1	0.717703349
14	0.556390977	0.986666667	0.711538462
15	0.563909774	1	0.721153846
16	0.559701493	1	0.717703349
17	0.559701493	1	0.717703349
18	0.559701493	1	0.717703349
19	0.559701493	1	0.717703349
20	0.563909774	1	0.721153846
21	0.559701493	1	0.717703349
22	0.556390977	0.986666667	0.711538462
23	0.559701493	1	0.717703349
24	0.559701493	1	0.717703349
25	0.559701493	1	0.717703349
26	0.559701493	1	0.717703349
27	0.559701493	1	0.717703349
28	0.556390977	0.986666667	0.711538462
29	0.559701493	1	0.717703349
30	0.559701493	1	0.717703349
31	0.559701493	1	0.717703349
32	0.556390977	0.986666667	0.711538462
33	0.559701493	1	0.717703349
34	0.559701493	1	0.717703349
35	0.556390977	0.986666667	0.711538462
36	0.563909774	1	0.721153846
37	0.556390977	0.986666667	0.711538462
38	0.559701493	1	0.717703349
39	0.559701493	1	0.717703349
40	0.559701493	1	0.717703349
41	0.559701493	1	0.717703349
42	0.559701493	1	0.717703349

43	0.559701493	1	0.717703349
44	0.559701493	1	0.717703349
45	0.559701493	1	0.717703349
46	0.559701493	1	0.717703349
47	0.559701493	1	0.717703349
48	0.559701493	1	0.717703349
49	0.563909774	1	0.721153846
50	0.559701493	1	0.717703349
51	0.559701493	1	0.717703349
52	0.559701493	1	0.717703349
53	0.559701493	1	0.717703349
54	0.559701493	1	0.717703349
55	0.559701493	1	0.717703349
56	0.556390977	0.986666667	0.711538462
57	0.559701493	1	0.717703349
58	0.559701493	1	0.717703349
59	0.559701493	1	0.717703349
60	0.559701493	1	0.717703349
61	0.559701493	1	0.717703349
62	0.559701493	1	0.717703349
63	0.559701493	1	0.717703349
64	0.559701493	1	0.717703349
65	0.559701493	1	0.717703349
66	0.559701493	1	0.717703349
67	0.559701493	1	0.717703349
68	0.559701493	1	0.717703349
69	0.559701493	1	0.717703349
70	0.559701493	1	0.717703349
71	0.563909774	1	0.721153846
72	0.559701493	1	0.717703349
73	0.559701493	1	0.717703349
74	0.559701493	1	0.717703349
75	0.559701493	1	0.717703349
76	0.559701493	1	0.717703349
77	0.560606061	0.986666667	0.714975845
78	0.559701493	1	0.717703349
79	0.559701493	1	0.717703349
80	0.559701493	1	0.717703349
81	0.559701493	1	0.717703349
82	0.559701493	1	0.717703349
83	0.559701493	1	0.717703349
84	0.559701493	1	0.717703349

Table G-11. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.440298507	1	0.611398964
1	0.333333333	0.016949153	0.032258065
9	0.285714286	0.033898305	0.060606061
11	1	0.033898305	0.06557377
12	0.5	0.016949153	0.032786885
15	1	0.016949153	0.033333333
20	1	0.016949153	0.033333333
36	1	0.016949153	0.033333333
49	1	0.016949153	0.033333333
71	1	0.016949153	0.033333333
77	0.5	0.016949153	0.032786885

Table G-12. P, R, F-Score for Females

7. Extracted Test Data: 20s and 40s

Male	Precision	Recall	F-Score
Baseline	0.558139535	1	0.71641791
1	0.558139535	1	0.71641791
2	0.5625	1	0.72
3	0.558139535	1	0.71641791
4	0.558139535	1	0.71641791
5	0.558139535	1	0.71641791
6	0.558139535	1	0.71641791
7	0.558139535	1	0.71641791
8	0.558139535	1	0.71641791
9	0.558139535	1	0.71641791
10	0.5546875	0.986111111	0.71
11	0.558139535	1	0.71641791
12	0.559055118	0.986111111	0.713567839
13	0.558139535	1	0.71641791
14	0.558139535	1	0.71641791
15	0.5625	1	0.72
16	0.5625	1	0.72
17	0.558139535	1	0.71641791
18	0.558139535	1	0.71641791
19	0.558139535	1	0.71641791
20	0.558139535	1	0.71641791
21	0.558139535	1	0.71641791
22	0.551181102	0.972222222	0.703517588
23	0.558139535	1	0.71641791
24	0.558139535	1	0.71641791
25	0.558139535	1	0.71641791
26	0.558139535	1	0.71641791
27	0.5546875	0.986111111	0.71
28	0.558139535	1	0.71641791
29	0.558139535	1	0.71641791
30	0.558139535	1	0.71641791
31	0.558139535	1	0.71641791
32	0.5546875	0.986111111	0.71
33	0.558139535	1	0.71641791
34	0.558139535	1	0.71641791
35	0.558139535	1	0.71641791
36	0.5625	1	0.72
37	0.558139535	1	0.71641791
38	0.558139535	1	0.71641791
39	0.558139535	1	0.71641791
40	0.558139535	1	0.71641791
41	0.558139535	1	0.71641791
42	0.558139535	1	0.71641791

43	0.558139535	1	0.71641791
44	0.558139535	1	0.71641791
45	0.558139535	1	0.71641791
46	0.558139535	1	0.71641791
47	0.558139535	1	0.71641791
48	0.558139535	1	0.71641791
49	0.558139535	1	0.71641791
50	0.558139535	1	0.71641791
51	0.558139535	1	0.71641791
52	0.558139535	1	0.71641791
53	0.558139535	1	0.71641791
54	0.558139535	1	0.71641791
55	0.558139535	1	0.71641791
56	0.558139535	1	0.71641791
57	0.558139535	1	0.71641791
58	0.558139535	1	0.71641791
59	0.558139535	1	0.71641791
60	0.5546875	0.986111111	0.71
61	0.558139535	1	0.71641791
62	0.558139535	1	0.71641791
63	0.558139535	1	0.71641791
64	0.558139535	1	0.71641791
65	0.558139535	1	0.71641791
66	0.558139535	1	0.71641791
67	0.558139535	1	0.71641791
68	0.558139535	1	0.71641791
69	0.558139535	1	0.71641791
70	0.558139535	1	0.71641791
71	0.558139535	1	0.71641791
72	0.558139535	1	0.71641791
73	0.5546875	0.986111111	0.71
74	0.558139535	1	0.71641791
75	0.558139535	1	0.71641791
76	0.558139535	1	0.71641791
77	0.558139535	1	0.71641791
78	0.558139535	1	0.71641791
79	0.558139535	1	0.71641791
80	0.558139535	1	0.71641791
81	0.558139535	1	0.71641791
82	0.558139535	1	0.71641791
83	0.558139535	1	0.71641791
84	0.558139535	1	0.71641791

Table G-13. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.441860465	1	0.612903226
2	1	0.01754386	0.034482759
12	0.5	0.01754386	0.033898305
15	1	0.01754386	0.034482759
16	1	0.01754386	0.034482759
36	1	0.01754386	0.034482759

Table G-14. P, R, F-Score for Females

8. Extracted Test Data: 20s and 50s

Male	Precision	Recall	F-Score
Baseline	0.560747664	1	0.718562874
1	0.560747664	1	0.718562874
2	0.566037736	1	0.722891566
3	0.560747664	1	0.718562874
4	0.560747664	1	0.718562874
5	0.560747664	1	0.718562874
6	0.560747664	1	0.718562874
7	0.560747664	1	0.718562874
8	0.560747664	1	0.718562874
9	0.554455446	0.933333333	0.695652174
10	0.556603774	0.983333333	0.710843373
11	0.567307692	0.983333333	0.719512195
12	0.561904762	0.983333333	0.715151515
13	0.560747664	1	0.718562874
14	0.560747664	1	0.718562874
15	0.566037736	1	0.722891566
16	0.560747664	1	0.718562874
17	0.560747664	1	0.718562874
18	0.560747664	1	0.718562874
19	0.560747664	1	0.718562874
20	0.560747664	1	0.718562874
21	0.560747664	1	0.718562874
22	0.556603774	0.983333333	0.710843373
23	0.560747664	1	0.718562874
24	0.560747664	1	0.718562874
25	0.560747664	1	0.718562874
26	0.560747664	1	0.718562874
27	0.556603774	0.983333333	0.710843373
28	0.560747664	1	0.718562874
29	0.560747664	1	0.718562874
30	0.560747664	1	0.718562874
31	0.560747664	1	0.718562874
32	0.556603774	0.983333333	0.710843373
33	0.560747664	1	0.718562874
34	0.560747664	1	0.718562874
35	0.560747664	1	0.718562874
36	0.560747664	1	0.718562874
37	0.560747664	1	0.718562874
38	0.560747664	1	0.718562874
39	0.560747664	1	0.718562874
40	0.556603774	0.983333333	0.710843373
41	0.560747664	1	0.718562874
42	0.560747664	1	0.718562874

43	0.560747664	1	0.718562874
44	0.560747664	1	0.718562874
45	0.560747664	1	0.718562874
46	0.560747664	1	0.718562874
47	0.560747664	1	0.718562874
48	0.560747664	1	0.718562874
49	0.560747664	1	0.718562874
50	0.560747664	1	0.718562874
51	0.560747664	1	0.718562874
52	0.560747664	1	0.718562874
53	0.556603774	0.983333333	0.710843373
54	0.560747664	1	0.718562874
55	0.556603774	0.983333333	0.710843373
56	0.560747664	1	0.718562874
57	0.560747664	1	0.718562874
58	0.560747664	1	0.718562874
59	0.560747664	1	0.718562874
60	0.560747664	1	0.718562874
61	0.560747664	1	0.718562874
62	0.560747664	1	0.718562874
63	0.560747664	1	0.718562874
64	0.560747664	1	0.718562874
65	0.560747664	1	0.718562874
66	0.560747664	1	0.718562874
67	0.560747664	1	0.718562874
68	0.560747664	1	0.718562874
69	0.560747664	1	0.718562874
70	0.560747664	1	0.718562874
71	0.560747664	1	0.718562874
72	0.560747664	1	0.718562874
73	0.560747664	1	0.718562874
74	0.560747664	1	0.718562874
75	0.560747664	1	0.718562874
76	0.560747664	1	0.718562874
77	0.560747664	1	0.718562874
78	0.560747664	1	0.718562874
79	0.560747664	1	0.718562874
80	0.560747664	1	0.718562874
81	0.560747664	1	0.718562874
82	0.560747664	1	0.718562874
83	0.560747664	1	0.718562874
84	0.560747664	1	0.718562874

Table G-15. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.439252336	1	0.61038961
2	1	0.021276596	0.041666667
9	0.333333333	0.042553191	0.075471698
11	0.666666667	0.042553191	0.08
12	0.5	0.021276596	0.040816327
15	1	0.021276596	0.041666667

Table G-16. P, R, F-Score for Females

9. Extracted Test Data: 30s and 40s

Male	Precision	Recall	F-Score
Baseline	0.623188406	1	0.767857143
1	0.623188406	1	0.767857143
2	0.611940299	0.953488372	0.745454545
3	0.626865672	0.976744186	0.763636364
4	0.632352941	1	0.774774775
5	0.623188406	1	0.767857143
6	0.623188406	1	0.767857143
7	0.623188406	1	0.767857143
8	0.623188406	1	0.767857143
9	0.623188406	1	0.767857143
10	0.626865672	0.976744186	0.763636364
11	0.623188406	1	0.767857143
12	0.623188406	1	0.767857143
13	0.617647059	0.976744186	0.756756757
14	0.617647059	0.976744186	0.756756757
15	0.617647059	0.976744186	0.756756757
16	0.623188406	1	0.767857143
17	0.623188406	1	0.767857143
18	0.623188406	1	0.767857143
19	0.623188406	1	0.767857143
20	0.623188406	1	0.767857143
21	0.623188406	1	0.767857143
22	0.617647059	0.976744186	0.756756757
23	0.623188406	1	0.767857143
24	0.623188406	1	0.767857143
25	0.623188406	1	0.767857143
26	0.623188406	1	0.767857143
27	0.623188406	1	0.767857143
28	0.623188406	1	0.767857143
29	0.623188406	1	0.767857143
30	0.623188406	1	0.767857143
31	0.623188406	1	0.767857143
32	0.623188406	1	0.767857143
33	0.623188406	1	0.767857143
34	0.623188406	1	0.767857143
35	0.623188406	1	0.767857143
36	0.623188406	1	0.767857143
37	0.623188406	1	0.767857143
38	0.623188406	1	0.767857143
39	0.623188406	1	0.767857143
40	0.617647059	0.976744186	0.756756757

41	0.623188406	1	0.767857143
42	0.623188406	1	0.767857143
43	0.623188406	1	0.767857143
44	0.623188406	1	0.767857143
45	0.632352941	1	0.774774775
46	0.623188406	1	0.767857143
47	0.623188406	1	0.767857143
48	0.623188406	1	0.767857143
49	0.623188406	1	0.767857143
50	0.623188406	1	0.767857143
51	0.623188406	1	0.767857143
52	0.623188406	1	0.767857143
53	0.623188406	1	0.767857143
54	0.623188406	1	0.767857143
55	0.626865672	0.976744186	0.763636364
56	0.617647059	0.976744186	0.756756757
57	0.623188406	1	0.767857143
58	0.623188406	1	0.767857143
59	0.617647059	0.976744186	0.756756757
60	0.617647059	0.976744186	0.756756757
61	0.617647059	0.976744186	0.756756757
62	0.623188406	1	0.767857143
63	0.623188406	1	0.767857143
64	0.623188406	1	0.767857143
65	0.623188406	1	0.767857143
66	0.623188406	1	0.767857143
67	0.623188406	1	0.767857143
68	0.623188406	1	0.767857143
69	0.623188406	1	0.767857143
70	0.623188406	1	0.767857143
71	0.623188406	1	0.767857143
72	0.623188406	1	0.767857143
73	0.623188406	1	0.767857143
74	0.623188406	1	0.767857143
75	0.623188406	1	0.767857143
76	0.623188406	1	0.767857143
77	0.623188406	1	0.767857143
78	0.623188406	1	0.767857143
79	0.623188406	1	0.767857143
80	0.623188406	1	0.767857143
81	0.623188406	1	0.767857143
82	0.623188406	1	0.767857143
83	0.623188406	1	0.767857143
84	0.623188406	1	0.767857143

Table G-17. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.376811594	1	0.547368421
3	0.5	0.038461538	0.071428571
4	1	0.038461538	0.074074074
10	0.5	0.038461538	0.071428571
45	1	0.038461538	0.074074074
55	0.5	0.038461538	0.071428571

Table G-18. P, R, F-Score for Females

10. Extracted Test Data: 30s and 50s

Male	Precision	Recall	F-Score
Baseline	0.659574468	1	0.794871795
1	0.659574468	1	0.794871795
2	0.673913043	1	0.805194805
3	0.652173913	0.967741935	0.779220779
4	0.659574468	1	0.794871795
5	0.659574468	1	0.794871795
6	0.659574468	1	0.794871795
7	0.659574468	1	0.794871795
8	0.659574468	1	0.794871795
9	0.666666667	0.967741935	0.789473684
10	0.659574468	1	0.794871795
11	0.659574468	1	0.794871795
12	0.659574468	1	0.794871795
13	0.659574468	1	0.794871795
14	0.652173913	0.967741935	0.779220779
15	0.659574468	1	0.794871795
16	0.659574468	1	0.794871795
17	0.659574468	1	0.794871795
18	0.659574468	1	0.794871795
19	0.659574468	1	0.794871795
20	0.659574468	1	0.794871795
21	0.659574468	1	0.794871795
22	0.659574468	1	0.794871795
23	0.659574468	1	0.794871795
24	0.659574468	1	0.794871795
25	0.659574468	1	0.794871795
26	0.659574468	1	0.794871795
27	0.659574468	1	0.794871795
28	0.659574468	1	0.794871795
29	0.659574468	1	0.794871795
30	0.659574468	1	0.794871795
31	0.659574468	1	0.794871795
32	0.659574468	1	0.794871795
33	0.659574468	1	0.794871795
34	0.659574468	1	0.794871795
35	0.652173913	0.967741935	0.779220779
36	0.659574468	1	0.794871795
37	0.652173913	0.967741935	0.779220779
38	0.659574468	1	0.794871795
39	0.659574468	1	0.794871795
40	0.659574468	1	0.794871795
41	0.659574468	1	0.794871795
42	0.659574468	1	0.794871795

43	0.659574468	1	0.794871795
44	0.659574468	1	0.794871795
45	0.659574468	1	0.794871795
46	0.659574468	1	0.794871795
47	0.673913043	1	0.805194805
48	0.673913043	1	0.805194805
49	0.673913043	1	0.805194805
50	0.659574468	1	0.794871795
51	0.659574468	1	0.794871795
52	0.659574468	1	0.794871795
53	0.659574468	1	0.794871795
54	0.652173913	0.967741935	0.779220779
55	0.659574468	1	0.794871795
56	0.652173913	0.967741935	0.779220779
57	0.659574468	1	0.794871795
58	0.659574468	1	0.794871795
59	0.659574468	1	0.794871795
60	0.659574468	1	0.794871795
61	0.659574468	1	0.794871795
62	0.659574468	1	0.794871795
63	0.659574468	1	0.794871795
64	0.659574468	1	0.794871795
65	0.659574468	1	0.794871795
66	0.659574468	1	0.794871795
67	0.659574468	1	0.794871795
68	0.659574468	1	0.794871795
69	0.659574468	1	0.794871795
70	0.659574468	1	0.794871795
71	0.659574468	1	0.794871795
72	0.659574468	1	0.794871795
73	0.659574468	1	0.794871795
74	0.659574468	1	0.794871795
75	0.659574468	1	0.794871795
76	0.659574468	1	0.794871795
77	0.659574468	1	0.794871795
78	0.659574468	1	0.794871795
79	0.659574468	1	0.794871795
80	0.659574468	1	0.794871795
81	0.659574468	1	0.794871795
82	0.659574468	1	0.794871795
83	0.673913043	1	0.805194805
84	0.659574468	1	0.794871795

Table G-19. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.340425532	1	0.507936508
2	1	0.0625	0.117647059
9	0.5	0.0625	0.111111111
47	1	0.0625	0.117647059
48	1	0.0625	0.117647059
49	1	0.0625	0.117647059
83	1	0.0625	0.117647059

Table G-20. P, R, F-Score for Females

11. Extracted Test Data: 40s and 50s

Male	Precision	Recall	F-Score
Baseline	0.666666667	1	0.8
1	0.666666667	1	0.8
2	0.666666667	1	0.8
3	0.666666667	1	0.8
4	0.682926829	1	0.811594203
5	0.666666667	1	0.8
6	0.666666667	1	0.8
7	0.666666667	1	0.8
8	0.666666667	1	0.8
9	0.65	0.928571429	0.764705882
10	0.658536585	0.964285714	0.782608696
11	0.666666667	1	0.8
12	0.666666667	1	0.8
13	0.666666667	1	0.8
14	0.666666667	1	0.8
15	0.666666667	1	0.8
16	0.666666667	1	0.8
17	0.666666667	1	0.8
18	0.666666667	1	0.8
19	0.666666667	1	0.8
20	0.666666667	1	0.8
21	0.666666667	1	0.8
22	0.666666667	1	0.8
23	0.666666667	1	0.8
24	0.666666667	1	0.8
25	0.666666667	1	0.8
26	0.666666667	1	0.8
27	0.666666667	1	0.8
28	0.666666667	1	0.8
29	0.666666667	1	0.8
30	0.666666667	1	0.8
31	0.666666667	1	0.8
32	0.666666667	1	0.8
33	0.666666667	1	0.8
34	0.666666667	1	0.8
35	0.666666667	1	0.8
36	0.666666667	1	0.8
37	0.666666667	1	0.8
38	0.666666667	1	0.8
39	0.666666667	1	0.8
40	0.666666667	1	0.8

41	0.666666667	1	0.8
42	0.666666667	1	0.8
43	0.666666667	1	0.8
44	0.666666667	1	0.8
45	0.682926829	1	0.811594203
46	0.682926829	1	0.811594203
47	0.666666667	1	0.8
48	0.666666667	1	0.8
49	0.666666667	1	0.8
50	0.666666667	1	0.8
51	0.658536585	0.964285714	0.782608696
52	0.666666667	1	0.8
53	0.666666667	1	0.8
54	0.666666667	1	0.8
55	0.666666667	1	0.8
56	0.666666667	1	0.8
57	0.666666667	1	0.8
58	0.666666667	1	0.8
59	0.666666667	1	0.8
60	0.658536585	0.964285714	0.782608696
61	0.666666667	1	0.8
62	0.666666667	1	0.8
63	0.666666667	1	0.8
64	0.666666667	1	0.8
65	0.666666667	1	0.8
66	0.666666667	1	0.8
67	0.666666667	1	0.8
68	0.666666667	1	0.8
69	0.666666667	1	0.8
70	0.675	0.964285714	0.794117647
71	0.666666667	1	0.8
72	0.666666667	1	0.8
73	0.666666667	1	0.8
74	0.666666667	1	0.8
75	0.666666667	1	0.8
76	0.666666667	1	0.8
77	0.666666667	1	0.8
78	0.666666667	1	0.8
79	0.666666667	1	0.8
80	0.666666667	1	0.8
81	0.666666667	1	0.8
82	0.666666667	1	0.8
83	0.666666667	1	0.8
84	0.666666667	1	0.8

Table G-21. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.333333333	1	0.5
4	1	0.071428571	0.133333333
45	1	0.071428571	0.133333333
46	1	0.071428571	0.133333333
70	0.5	0.071428571	0.125

Table G-22. P, R, F-Score for Females

12. Extracted Test Data: Under 26 and 26 or Over

Male	Precision	Recall	F-Score
Baseline	0.512295082	1	0.677506775
1	0.5125	0.984	0.673972603
2	0.522123894	0.944	0.672364672
3	0.510288066	0.992	0.673913043
4	0.510288066	0.992	0.673913043
5	0.512295082	1	0.677506775
6	0.512295082	1	0.677506775
7	0.512295082	1	0.677506775
8	0.512295082	1	0.677506775
9	0.514403292	1	0.679347826
10	0.514893617	0.968	0.672222222
11	0.510288066	0.992	0.673913043
12	0.512396694	0.992	0.675749319
13	0.512295082	1	0.677506775
14	0.510288066	0.992	0.673913043
15	0.512295082	1	0.677506775
16	0.516528926	1	0.68119891
17	0.512295082	1	0.677506775
18	0.518672199	1	0.683060109
19	0.512295082	1	0.677506775
20	0.514403292	1	0.679347826
21	0.512295082	1	0.677506775
22	0.510288066	0.992	0.673913043
23	0.512295082	1	0.677506775
24	0.512295082	1	0.677506775
25	0.512295082	1	0.677506775
26	0.514403292	1	0.679347826
27	0.512295082	1	0.677506775
28	0.510288066	0.992	0.673913043
29	0.512295082	1	0.677506775
30	0.512295082	1	0.677506775
31	0.512295082	1	0.677506775
32	0.512295082	1	0.677506775
33	0.512295082	1	0.677506775
34	0.512295082	1	0.677506775
35	0.514522822	0.992	0.677595628
36	0.512295082	1	0.677506775
37	0.514403292	1	0.679347826
38	0.514403292	1	0.679347826
39	0.512295082	1	0.677506775
40	0.512295082	1	0.677506775
41	0.512295082	1	0.677506775
42	0.512295082	1	0.677506775

43	0.512295082	1	0.677506775
44	0.512295082	1	0.677506775
45	0.512295082	1	0.677506775
46	0.512295082	1	0.677506775
47	0.514403292	1	0.679347826
48	0.512295082	1	0.677506775
49	0.512295082	1	0.677506775
50	0.512295082	1	0.677506775
51	0.512295082	1	0.677506775
52	0.512295082	1	0.677506775
53	0.512295082	1	0.677506775
54	0.512295082	1	0.677506775
55	0.512295082	1	0.677506775
56	0.521008403	0.992	0.683195592
57	0.512295082	1	0.677506775
58	0.512295082	1	0.677506775
59	0.512295082	1	0.677506775
60	0.510288066	0.992	0.673913043
61	0.512295082	1	0.677506775
62	0.512295082	1	0.677506775
63	0.512295082	1	0.677506775
64	0.512295082	1	0.677506775
65	0.512295082	1	0.677506775
66	0.512295082	1	0.677506775
67	0.512295082	1	0.677506775
68	0.512295082	1	0.677506775
69	0.512295082	1	0.677506775
70	0.512295082	1	0.677506775
71	0.512295082	1	0.677506775
72	0.512295082	1	0.677506775
73	0.512295082	1	0.677506775
74	0.512295082	1	0.677506775
75	0.512295082	1	0.677506775
76	0.512295082	1	0.677506775
77	0.416666667	0.2	0.27027027
78	0.508403361	0.968	0.666666667
79	0.512295082	1	0.677506775
80	0.512295082	1	0.677506775
81	0.512295082	1	0.677506775
82	0.512295082	1	0.677506775
83	0.512295082	1	0.677506775
84	0.512295082	1	0.677506775

Table G-23. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.487704918	1	0.655647383
1	0.5	0.016806723	0.032520325
2	0.611111111	0.092436975	0.160583942
9	1	0.008403361	0.016666667
10	0.555555556	0.042016807	0.078125
12	0.5	0.008403361	0.016528926
16	1	0.016806723	0.033057851
18	1	0.025210084	0.049180328
20	1	0.008403361	0.016666667
26	1	0.008403361	0.016666667
35	0.666666667	0.016806723	0.032786885
37	1	0.008403361	0.016666667
38	1	0.008403361	0.016666667
47	1	0.008403361	0.016666667
56	0.833333333	0.042016807	0.08
77	0.456521739	0.705882353	0.554455446
78	0.333333333	0.016806723	0.032

Table G-24. P, R, F-Score for Females

B. AGE: MULTI-CLASS (5-WAY) CLASSIFICATION WITH PRIOR

1. All Test Data

13-19	Precision	Recall	F-Score
Baseline	0.278688525	1	0.435897436
14	0.4	0.029411765	0.054794521
23	1	0.014705882	0.028985507
26	1	0.014705882	0.028985507
29	1	0.014705882	0.028985507
33	0.5	0.014705882	0.028571429
34	1	0.014705882	0.028985507
35	0.5	0.014705882	0.028571429
37	0.5	0.014705882	0.028571429
38	0.5	0.014705882	0.028571429
47	1	0.014705882	0.028985507
57	1	0.014705882	0.028985507
69	1	0.014705882	0.028985507
70	1	0.014705882	0.028985507
77	0.571428571	0.058823529	0.106666667

Table G-25. P, R, F-Score for Teens

20-29	Precision	Recall	F-Score
Baseline	0.397540984	1	0.568914956
1	0.397540984	1	0.568914956
2	0.397540984	1	0.568914956
3	0.399176955	1	0.570588235
4	0.399176955	1	0.570588235
5	0.397540984	1	0.568914956
6	0.397540984	1	0.568914956
7	0.397540984	1	0.568914956
8	0.397540984	1	0.568914956
9	0.397540984	1	0.568914956
10	0.397540984	1	0.568914956
11	0.395061728	0.989690722	0.564705882
12	0.392561983	0.979381443	0.560471976
13	0.397540984	1	0.568914956
14	0.40167364	0.989690722	0.571428571
15	0.395061728	0.989690722	0.564705882
16	0.397540984	1	0.568914956
17	0.395061728	0.989690722	0.564705882
18	0.397540984	1	0.568914956
19	0.397540984	1	0.568914956
20	0.397540984	1	0.568914956
21	0.397540984	1	0.568914956
22	0.395061728	0.989690722	0.564705882
23	0.399176955	1	0.570588235
24	0.397540984	1	0.568914956
25	0.397540984	1	0.568914956
26	0.399176955	1	0.570588235
27	0.395061728	0.989690722	0.564705882
28	0.395061728	0.989690722	0.564705882
29	0.399176955	1	0.570588235
30	0.397540984	1	0.568914956
31	0.395061728	0.989690722	0.564705882
32	0.395061728	0.989690722	0.564705882
33	0.396694215	0.989690722	0.566371681
34	0.399176955	1	0.570588235
35	0.402489627	1	0.573964497
36	0.397540984	1	0.568914956
37	0.402489627	1	0.573964497
38	0.402489627	1	0.573964497
39	0.397540984	1	0.568914956
40	0.399176955	1	0.570588235
41	0.397540984	1	0.568914956
42	0.399176955	1	0.570588235

43	0.397540984	1	0.568914956
44	0.397540984	1	0.568914956
45	0.397540984	1	0.568914956
46	0.397540984	1	0.568914956
47	0.399176955	1	0.570588235
48	0.397540984	1	0.568914956
49	0.397540984	1	0.568914956
50	0.397540984	1	0.568914956
51	0.397540984	1	0.568914956
52	0.397540984	1	0.568914956
53	0.395061728	0.989690722	0.564705882
54	0.397540984	1	0.568914956
55	0.396694215	0.989690722	0.566371681
56	0.399176955	1	0.570588235
57	0.399176955	1	0.570588235
58	0.397540984	1	0.568914956
59	0.397540984	1	0.568914956
60	0.397540984	1	0.568914956
61	0.397540984	1	0.568914956
62	0.386554622	0.948453608	0.549253731
63	0.399176955	1	0.570588235
64	0.397540984	1	0.568914956
65	0.397540984	1	0.568914956
66	0.397540984	1	0.568914956
67	0.397540984	1	0.568914956
68	0.397540984	1	0.568914956
69	0.399176955	1	0.570588235
70	0.399176955	1	0.570588235
71	0.399176955	1	0.570588235
72	0.397540984	1	0.568914956
73	0.395061728	0.989690722	0.564705882
74	0.397540984	1	0.568914956
75	0.397540984	1	0.568914956
76	0.397540984	1	0.568914956
77	0.400843882	0.979381443	0.568862275
78	0.397540984	1	0.568914956
79	0.397540984	1	0.568914956
80	0.397540984	1	0.568914956
81	0.397540984	1	0.568914956
82	0.397540984	1	0.568914956
83	0.397540984	1	0.568914956
84	0.397540984	1	0.568914956

Table G-26. P, R, F-Score for Males

C. AGE: BINARY CLASSIFICATION WITH PRIOR

1. Extracted Test Data: Teens and 20s

13-19	Precision	Recall	F-Score
Baseline	0.412121212	1	0.583690987
14	0.666666667	0.029411765	0.056338028
17	0.666666667	0.029411765	0.056338028
23	1	0.014705882	0.028985507
29	1	0.014705882	0.028985507
34	1	0.014705882	0.028985507
35	1	0.029411765	0.057142857
37	1	0.014705882	0.028985507
38	1	0.014705882	0.028985507
42	1	0.014705882	0.028985507
47	1	0.014705882	0.028985507
55	0.5	0.014705882	0.028571429
57	1	0.014705882	0.028985507
63	1	0.014705882	0.028985507
69	1	0.014705882	0.028985507
70	1	0.014705882	0.028985507
71	1	0.014705882	0.028985507
77	0.857142857	0.088235294	0.16

Table G-27.

P, R, F-Score for Teens

20-29	Precision	Recall	F-Score
Baseline	0.587878788	1	0.740458015
1	0.587878788	1	0.740458015
2	0.587878788	1	0.740458015
3	0.587878788	1	0.740458015
4	0.587878788	1	0.740458015
5	0.587878788	1	0.740458015
6	0.587878788	1	0.740458015
7	0.587878788	1	0.740458015
8	0.587878788	1	0.740458015
9	0.587878788	1	0.740458015
10	0.587878788	1	0.740458015
11	0.585365854	0.989690722	0.735632184
12	0.587878788	1	0.740458015
13	0.587878788	1	0.740458015
14	0.592592593	0.989690722	0.741312741
15	0.585365854	0.989690722	0.735632184
16	0.587878788	1	0.740458015
17	0.592592593	0.989690722	0.741312741
18	0.587878788	1	0.740458015
19	0.587878788	1	0.740458015
20	0.587878788	1	0.740458015
21	0.587878788	1	0.740458015
22	0.585365854	0.989690722	0.735632184
23	0.591463415	1	0.743295019
24	0.587878788	1	0.740458015
25	0.587878788	1	0.740458015
26	0.587878788	1	0.740458015
27	0.585365854	0.989690722	0.735632184
28	0.585365854	0.989690722	0.735632184
29	0.591463415	1	0.743295019
30	0.587878788	1	0.740458015
31	0.587878788	1	0.740458015
32	0.587878788	1	0.740458015
33	0.585365854	0.989690722	0.735632184
34	0.591463415	1	0.743295019
35	0.595092025	1	0.746153846
36	0.587878788	1	0.740458015
37	0.591463415	1	0.743295019
38	0.591463415	1	0.743295019
39	0.587878788	1	0.740458015
40	0.587878788	1	0.740458015
41	0.587878788	1	0.740458015
42	0.591463415	1	0.743295019

43	0.587878788	1	0.740458015
44	0.587878788	1	0.740458015
45	0.587878788	1	0.740458015
46	0.587878788	1	0.740458015
47	0.591463415	1	0.743295019
48	0.587878788	1	0.740458015
49	0.587878788	1	0.740458015
50	0.587878788	1	0.740458015
51	0.587878788	1	0.740458015
52	0.587878788	1	0.740458015
53	0.585365854	0.989690722	0.735632184
54	0.587878788	1	0.740458015
55	0.588957055	0.989690722	0.738461538
56	0.587878788	1	0.740458015
57	0.591463415	1	0.743295019
58	0.587878788	1	0.740458015
59	0.587878788	1	0.740458015
60	0.587878788	1	0.740458015
61	0.587878788	1	0.740458015
62	0.575	0.948453608	0.715953307
63	0.591463415	1	0.743295019
64	0.587878788	1	0.740458015
65	0.587878788	1	0.740458015
66	0.587878788	1	0.740458015
67	0.587878788	1	0.740458015
68	0.587878788	1	0.740458015
69	0.591463415	1	0.743295019
70	0.591463415	1	0.743295019
71	0.591463415	1	0.743295019
72	0.587878788	1	0.740458015
73	0.587878788	1	0.740458015
74	0.587878788	1	0.740458015
75	0.587878788	1	0.740458015
76	0.587878788	1	0.740458015
77	0.607594937	0.989690722	0.752941176
78	0.587878788	1	0.740458015
79	0.587878788	1	0.740458015
80	0.587878788	1	0.740458015
81	0.587878788	1	0.740458015
82	0.587878788	1	0.740458015
83	0.587878788	1	0.740458015
84	0.587878788	1	0.740458015

Table G-28. P, R, F-Score for 20s

2. Extracted Test Data: Teens and 30s

13-19	Precision	Recall	F-Score
Baseline	0.647619048	1	0.786127168
1	0.647619048	1	0.786127168
2	0.647619048	1	0.786127168
3	0.653846154	1	0.790697674
4	0.647619048	1	0.786127168
5	0.647619048	1	0.786127168
6	0.647619048	1	0.786127168
7	0.647619048	1	0.786127168
8	0.647619048	1	0.786127168
9	0.647619048	1	0.786127168
10	0.647619048	1	0.786127168
11	0.643564356	0.955882353	0.769230769
12	0.647058824	0.970588235	0.776470588
13	0.647619048	1	0.786127168
14	0.647619048	1	0.786127168
15	0.647619048	1	0.786127168
16	0.647619048	1	0.786127168
17	0.647619048	1	0.786127168
18	0.647619048	1	0.786127168
19	0.653846154	1	0.790697674
20	0.653846154	1	0.790697674
21	0.647619048	1	0.786127168
22	0.647619048	1	0.786127168
23	0.647619048	1	0.786127168
24	0.647619048	1	0.786127168
25	0.644230769	0.985294118	0.779069767
26	0.647619048	1	0.786127168
27	0.647619048	1	0.786127168
28	0.647619048	1	0.786127168
29	0.644230769	0.985294118	0.779069767
30	0.644230769	0.985294118	0.779069767
31	0.647619048	1	0.786127168
32	0.647619048	1	0.786127168
33	0.644230769	0.985294118	0.779069767
34	0.644230769	0.985294118	0.779069767
35	0.647619048	1	0.786127168
36	0.647619048	1	0.786127168
37	0.644230769	0.985294118	0.779069767
38	0.644230769	0.985294118	0.779069767
39	0.647619048	1	0.786127168
40	0.647619048	1	0.786127168
41	0.647619048	1	0.786127168
42	0.647619048	1	0.786127168

43	0.647619048	1	0.786127168
44	0.647619048	1	0.786127168
45	0.647619048	1	0.786127168
46	0.644230769	0.985294118	0.779069767
47	0.644230769	0.985294118	0.779069767
48	0.644230769	0.985294118	0.779069767
49	0.647619048	1	0.786127168
50	0.653846154	1	0.790697674
51	0.647619048	1	0.786127168
52	0.647619048	1	0.786127168
53	0.647619048	1	0.786127168
54	0.647619048	1	0.786127168
55	0.647619048	1	0.786127168
56	0.647619048	1	0.786127168
57	0.644230769	0.985294118	0.779069767
58	0.647619048	1	0.786127168
59	0.647619048	1	0.786127168
60	0.647619048	1	0.786127168
61	0.647619048	1	0.786127168
62	0.647619048	1	0.786127168
63	0.644230769	0.985294118	0.779069767
64	0.647619048	1	0.786127168
65	0.647619048	1	0.786127168
66	0.647619048	1	0.786127168
67	0.647619048	1	0.786127168
68	0.647619048	1	0.786127168
69	0.647619048	1	0.786127168
70	0.647619048	1	0.786127168
71	0.647619048	1	0.786127168
72	0.647619048	1	0.786127168
73	0.647619048	1	0.786127168
74	0.647619048	1	0.786127168
75	0.647619048	1	0.786127168
76	0.647619048	1	0.786127168
77	0.647619048	1	0.786127168
78	0.647619048	1	0.786127168
79	0.647619048	1	0.786127168
80	0.647619048	1	0.786127168
81	0.647619048	1	0.786127168
82	0.647619048	1	0.786127168
83	0.647619048	1	0.786127168
84	0.647619048	1	0.786127168

Table G-29. P, R, F-Score for Teens

30-39	Precision	Recall	F-Score
Baseline	0.352380952	1	0.521126761
3	1	0.027027027	0.052631579
11	0.25	0.027027027	0.048780488
12	0.333333333	0.027027027	0.05
19	1	0.027027027	0.052631579
20	1	0.027027027	0.052631579
50	1	0.027027027	0.052631579

Table G-30. P, R, F-Score for 30s

3. Extracted Test Data: Teens and 40s

13-19	Precision	Recall	F-Score
Baseline	0.68	1	0.80952381
1	0.68	1	0.80952381
2	0.68	1	0.80952381
3	0.68	1	0.80952381
4	0.68	1	0.80952381
5	0.68	1	0.80952381
6	0.68	1	0.80952381
7	0.68	1	0.80952381
8	0.68	1	0.80952381
9	0.68	1	0.80952381
10	0.68	1	0.80952381
11	0.68	1	0.80952381
12	0.68	1	0.80952381
13	0.68	1	0.80952381
14	0.68	1	0.80952381
15	0.68	1	0.80952381
16	0.68	1	0.80952381
17	0.68	1	0.80952381
18	0.68	1	0.80952381
19	0.68	1	0.80952381
20	0.68	1	0.80952381
21	0.68	1	0.80952381
22	0.68	1	0.80952381
23	0.676767677	0.985294118	0.80239521
24	0.676767677	0.985294118	0.80239521
25	0.676767677	0.985294118	0.80239521
26	0.676767677	0.985294118	0.80239521
27	0.68	1	0.80952381
28	0.68	1	0.80952381
29	0.676767677	0.985294118	0.80239521
30	0.676767677	0.985294118	0.80239521
31	0.68	1	0.80952381
32	0.68	1	0.80952381
33	0.676767677	0.985294118	0.80239521
34	0.676767677	0.985294118	0.80239521
35	0.676767677	0.985294118	0.80239521
36	0.68	1	0.80952381
37	0.676767677	0.985294118	0.80239521
38	0.68	1	0.80952381
39	0.68	1	0.80952381
40	0.686868687	1	0.814371257
41	0.68	1	0.80952381
42	0.68	1	0.80952381

43	0.68	1	0.80952381
44	0.68	1	0.80952381
45	0.68	1	0.80952381
46	0.68	1	0.80952381
47	0.68	1	0.80952381
48	0.68	1	0.80952381
49	0.68	1	0.80952381
50	0.68	1	0.80952381
51	0.68	1	0.80952381
52	0.68	1	0.80952381
53	0.68	1	0.80952381
54	0.68	1	0.80952381
55	0.68	1	0.80952381
56	0.68	1	0.80952381
57	0.68	1	0.80952381
58	0.68	1	0.80952381
59	0.68	1	0.80952381
60	0.68	1	0.80952381
61	0.68	1	0.80952381
62	0.68	1	0.80952381
63	0.676767677	0.985294118	0.80239521
64	0.68	1	0.80952381
65	0.68	1	0.80952381
66	0.68	1	0.80952381
67	0.68	1	0.80952381
68	0.686868687	1	0.814371257
69	0.68	1	0.80952381
70	0.68	1	0.80952381
71	0.676767677	0.985294118	0.80239521
72	0.68	1	0.80952381
73	0.68	1	0.80952381
74	0.68	1	0.80952381
75	0.68	1	0.80952381
76	0.68	1	0.80952381
77	0.68	1	0.80952381
78	0.68	1	0.80952381
79	0.68	1	0.80952381
80	0.68	1	0.80952381
81	0.68	1	0.80952381
82	0.68	1	0.80952381
83	0.68	1	0.80952381
84	0.68	1	0.80952381

Table G-31. P, R, F-Score for Teens

40-49	Precision	Recall	F-Score
Baseline	0.32	1	0.484848485
40	1	0.03125	0.060606061
68	1	0.03125	0.060606061

Table G-32.

P, R, F-Score for 40s

4. Extracted Test Data: Teens and 50s

13-19	Precision	Recall	F-Score
Baseline	0.871794872	1	0.931506849
1	0.871794872	1	0.931506849
2	0.871794872	1	0.931506849
3	0.871794872	1	0.931506849
4	0.871794872	1	0.931506849
5	0.871794872	1	0.931506849
6	0.871794872	1	0.931506849
7	0.871794872	1	0.931506849
8	0.871794872	1	0.931506849
9	0.87012987	0.985294118	0.924137931
10	0.871794872	1	0.931506849
11	0.871794872	1	0.931506849
12	0.871794872	1	0.931506849
13	0.871794872	1	0.931506849
14	0.871794872	1	0.931506849
15	0.871794872	1	0.931506849
16	0.871794872	1	0.931506849
17	0.871794872	1	0.931506849
18	0.871794872	1	0.931506849
19	0.871794872	1	0.931506849
20	0.871794872	1	0.931506849
21	0.871794872	1	0.931506849
22	0.871794872	1	0.931506849
23	0.871794872	1	0.931506849
24	0.871794872	1	0.931506849
25	0.87012987	0.985294118	0.924137931
26	0.87012987	0.985294118	0.924137931
27	0.871794872	1	0.931506849
28	0.871794872	1	0.931506849
29	0.87012987	0.985294118	0.924137931
30	0.87012987	0.985294118	0.924137931
31	0.871794872	1	0.931506849
32	0.871794872	1	0.931506849
33	0.87012987	0.985294118	0.924137931
34	0.87012987	0.985294118	0.924137931
35	0.87012987	0.985294118	0.924137931
36	0.871794872	1	0.931506849
37	0.87012987	0.985294118	0.924137931
38	0.87012987	0.985294118	0.924137931
39	0.871794872	1	0.931506849
40	0.871794872	1	0.931506849
41	0.871794872	1	0.931506849
42	0.871794872	1	0.931506849

43	0.871794872	1	0.931506849
44	0.871794872	1	0.931506849
45	0.871794872	1	0.931506849
46	0.871794872	1	0.931506849
47	0.87012987	0.985294118	0.924137931
48	0.87012987	0.985294118	0.924137931
49	0.871794872	1	0.931506849
50	0.871794872	1	0.931506849
51	0.871794872	1	0.931506849
52	0.871794872	1	0.931506849
53	0.871794872	1	0.931506849
54	0.871794872	1	0.931506849
55	0.871794872	1	0.931506849
56	0.871794872	1	0.931506849
57	0.87012987	0.985294118	0.924137931
58	0.871794872	1	0.931506849
59	0.871794872	1	0.931506849
60	0.871794872	1	0.931506849
61	0.871794872	1	0.931506849
62	0.871794872	1	0.931506849
63	0.87012987	0.985294118	0.924137931
64	0.871794872	1	0.931506849
65	0.871794872	1	0.931506849
66	0.871794872	1	0.931506849
67	0.871794872	1	0.931506849
68	0.871794872	1	0.931506849
69	0.871794872	1	0.931506849
70	0.871794872	1	0.931506849
71	0.87012987	0.985294118	0.924137931
72	0.871794872	1	0.931506849
73	0.871794872	1	0.931506849
74	0.871794872	1	0.931506849
75	0.871794872	1	0.931506849
76	0.871794872	1	0.931506849
77	0.871794872	1	0.931506849
78	0.871794872	1	0.931506849
79	0.871794872	1	0.931506849
80	0.871794872	1	0.931506849
81	0.871794872	1	0.931506849
82	0.871794872	1	0.931506849
83	0.871794872	1	0.931506849
84	0.871794872	1	0.931506849

Table G-33. P, R, F-Score for Teens

5. Extracted Test Data: 20s and 30s

20-29	Precision	Recall	F-Score
Baseline	0.723880597	1	0.83982684
1	0.723880597	1	0.83982684
2	0.723880597	1	0.83982684
3	0.729323308	1	0.843478261
4	0.729323308	1	0.843478261
5	0.723880597	1	0.83982684
6	0.723880597	1	0.83982684
7	0.723880597	1	0.83982684
8	0.723880597	1	0.83982684
9	0.723880597	1	0.83982684
10	0.723880597	1	0.83982684
11	0.723880597	1	0.83982684
12	0.723880597	1	0.83982684
13	0.723880597	1	0.83982684
14	0.729323308	1	0.843478261
15	0.721804511	0.989690722	0.834782609
16	0.723880597	1	0.83982684
17	0.721804511	0.989690722	0.834782609
18	0.723880597	1	0.83982684
19	0.723880597	1	0.83982684
20	0.723880597	1	0.83982684
21	0.723880597	1	0.83982684
22	0.721804511	0.989690722	0.834782609
23	0.723880597	1	0.83982684
24	0.723880597	1	0.83982684
25	0.723880597	1	0.83982684
26	0.723880597	1	0.83982684
27	0.723880597	1	0.83982684
28	0.721804511	0.989690722	0.834782609
29	0.723880597	1	0.83982684
30	0.723880597	1	0.83982684
31	0.723880597	1	0.83982684
32	0.723880597	1	0.83982684
33	0.723880597	1	0.83982684
34	0.723880597	1	0.83982684
35	0.729323308	1	0.843478261
36	0.721804511	0.989690722	0.834782609
37	0.729323308	1	0.843478261
38	0.729323308	1	0.843478261
39	0.723880597	1	0.83982684
40	0.723880597	1	0.83982684
41	0.723880597	1	0.83982684
42	0.723880597	1	0.83982684

43	0.723880597	1	0.83982684
44	0.723880597	1	0.83982684
45	0.723880597	1	0.83982684
46	0.723880597	1	0.83982684
47	0.723880597	1	0.83982684
48	0.729323308	1	0.843478261
49	0.723880597	1	0.83982684
50	0.723880597	1	0.83982684
51	0.723880597	1	0.83982684
52	0.723880597	1	0.83982684
53	0.723880597	1	0.83982684
54	0.723880597	1	0.83982684
55	0.723880597	1	0.83982684
56	0.723880597	1	0.83982684
57	0.723880597	1	0.83982684
58	0.723880597	1	0.83982684
59	0.723880597	1	0.83982684
60	0.723880597	1	0.83982684
61	0.721804511	0.989690722	0.834782609
62	0.723880597	1	0.83982684
63	0.723880597	1	0.83982684
64	0.723880597	1	0.83982684
65	0.723880597	1	0.83982684
66	0.723880597	1	0.83982684
67	0.723880597	1	0.83982684
68	0.723880597	1	0.83982684
69	0.723880597	1	0.83982684
70	0.723880597	1	0.83982684
71	0.729323308	1	0.843478261
72	0.723880597	1	0.83982684
73	0.723880597	1	0.83982684
74	0.723880597	1	0.83982684
75	0.723880597	1	0.83982684
76	0.723880597	1	0.83982684
77	0.723880597	1	0.83982684
78	0.723880597	1	0.83982684
79	0.723880597	1	0.83982684
80	0.723880597	1	0.83982684
81	0.723880597	1	0.83982684
82	0.723880597	1	0.83982684
83	0.723880597	1	0.83982684
84	0.723880597	1	0.83982684

Table G-34. P, R, F-Score for 20s

30-39	Precision	Recall	F-Score
Baseline	0.276119403	1	0.432748538
3	1	0.027027027	0.052631579
4	1	0.027027027	0.052631579
14	1	0.027027027	0.052631579
35	1	0.027027027	0.052631579
37	1	0.027027027	0.052631579
38	1	0.027027027	0.052631579
48	1	0.027027027	0.052631579
71	1	0.027027027	0.052631579

Table G-35.

P, R, F-Score for 30s

6. Extracted Test Data: 20s and 40s

20-29	Precision	Recall	F-Score
Baseline	0.751937984	1	0.85840708
1	0.751937984	1	0.85840708
2	0.751937984	1	0.85840708
3	0.751937984	1	0.85840708
4	0.751937984	1	0.85840708
5	0.751937984	1	0.85840708
6	0.751937984	1	0.85840708
7	0.751937984	1	0.85840708
8	0.751937984	1	0.85840708
9	0.751937984	1	0.85840708
10	0.751937984	1	0.85840708
11	0.751937984	1	0.85840708
12	0.751937984	1	0.85840708
13	0.751937984	1	0.85840708
14	0.75	0.989690722	0.853333333
15	0.75	0.989690722	0.853333333
16	0.751937984	1	0.85840708
17	0.75	0.989690722	0.853333333
18	0.751937984	1	0.85840708
19	0.751937984	1	0.85840708
20	0.751937984	1	0.85840708
21	0.751937984	1	0.85840708
22	0.755905512	0.989690722	0.857142857
23	0.751937984	1	0.85840708
24	0.751937984	1	0.85840708
25	0.751937984	1	0.85840708
26	0.7578125	1	0.862222222
27	0.75	0.989690722	0.853333333
28	0.748031496	0.979381443	0.848214286
29	0.751937984	1	0.85840708
30	0.751937984	1	0.85840708
31	0.75	0.989690722	0.853333333
32	0.751937984	1	0.85840708
33	0.75	0.989690722	0.853333333
34	0.75	0.989690722	0.853333333
35	0.751937984	1	0.85840708
36	0.75	0.989690722	0.853333333
37	0.751937984	1	0.85840708
38	0.751937984	1	0.85840708
39	0.751937984	1	0.85840708
40	0.751937984	1	0.85840708
41	0.751937984	1	0.85840708
42	0.751937984	1	0.85840708

43	0.751937984	1	0.85840708
44	0.751937984	1	0.85840708
45	0.751937984	1	0.85840708
46	0.751937984	1	0.85840708
47	0.751937984	1	0.85840708
48	0.751937984	1	0.85840708
49	0.751937984	1	0.85840708
50	0.7578125	1	0.862222222
51	0.751937984	1	0.85840708
52	0.751937984	1	0.85840708
53	0.751937984	1	0.85840708
54	0.751937984	1	0.85840708
55	0.751937984	1	0.85840708
56	0.751937984	1	0.85840708
57	0.751937984	1	0.85840708
58	0.751937984	1	0.85840708
59	0.751937984	1	0.85840708
60	0.751937984	1	0.85840708
61	0.75	0.989690722	0.853333333
62	0.751937984	1	0.85840708
63	0.751937984	1	0.85840708
64	0.751937984	1	0.85840708
65	0.751937984	1	0.85840708
66	0.751937984	1	0.85840708
67	0.751937984	1	0.85840708
68	0.751937984	1	0.85840708
69	0.751937984	1	0.85840708
70	0.751937984	1	0.85840708
71	0.751937984	1	0.85840708
72	0.751937984	1	0.85840708
73	0.75	0.989690722	0.853333333
74	0.751937984	1	0.85840708
75	0.751937984	1	0.85840708
76	0.751937984	1	0.85840708
77	0.751937984	1	0.85840708
78	0.751937984	1	0.85840708
79	0.751937984	1	0.85840708
80	0.751937984	1	0.85840708
81	0.751937984	1	0.85840708
82	0.751937984	1	0.85840708
83	0.751937984	1	0.85840708
84	0.751937984	1	0.85840708

Table G-36. P, R, F-Score for 20s

40-49	Precision	Recall	F-Score
Baseline	0.248062016	1	0.397515528
22	0.5	0.03125	0.058823529
26	1	0.03125	0.060606061
50	1	0.03125	0.060606061

Table G-37. P, R, F-Score for 40s

7. Extracted Test Data: 20s and 50s

20-29	Precision	Recall	F-Score
Baseline	0.906542056	1	0.950980392
1	0.906542056	1	0.950980392
2	0.906542056	1	0.950980392
3	0.906542056	1	0.950980392
4	0.906542056	1	0.950980392
5	0.906542056	1	0.950980392
6	0.906542056	1	0.950980392
7	0.906542056	1	0.950980392
8	0.906542056	1	0.950980392
9	0.906542056	1	0.950980392
10	0.906542056	1	0.950980392
11	0.906542056	1	0.950980392
12	0.906542056	1	0.950980392
13	0.906542056	1	0.950980392
14	0.905660377	0.989690722	0.945812808
15	0.905660377	0.989690722	0.945812808
16	0.906542056	1	0.950980392
17	0.906542056	1	0.950980392
18	0.906542056	1	0.950980392
19	0.906542056	1	0.950980392
20	0.906542056	1	0.950980392
21	0.906542056	1	0.950980392
22	0.905660377	0.989690722	0.945812808
23	0.906542056	1	0.950980392
24	0.906542056	1	0.950980392
25	0.906542056	1	0.950980392
26	0.906542056	1	0.950980392
27	0.905660377	0.989690722	0.945812808
28	0.904761905	0.979381443	0.940594059
29	0.906542056	1	0.950980392
30	0.906542056	1	0.950980392
31	0.906542056	1	0.950980392
32	0.906542056	1	0.950980392
33	0.906542056	1	0.950980392
34	0.905660377	0.989690722	0.945812808
35	0.906542056	1	0.950980392
36	0.906542056	1	0.950980392
37	0.906542056	1	0.950980392
38	0.906542056	1	0.950980392
39	0.906542056	1	0.950980392
40	0.906542056	1	0.950980392

41	0.906542056	1	0.950980392
42	0.906542056	1	0.950980392
43	0.906542056	1	0.950980392
44	0.906542056	1	0.950980392
45	0.906542056	1	0.950980392
46	0.906542056	1	0.950980392
47	0.906542056	1	0.950980392
48	0.906542056	1	0.950980392
49	0.906542056	1	0.950980392
50	0.906542056	1	0.950980392
51	0.906542056	1	0.950980392
52	0.906542056	1	0.950980392
53	0.906542056	1	0.950980392
54	0.906542056	1	0.950980392
55	0.906542056	1	0.950980392
56	0.906542056	1	0.950980392
57	0.906542056	1	0.950980392
58	0.906542056	1	0.950980392
59	0.906542056	1	0.950980392
60	0.906542056	1	0.950980392
61	0.905660377	0.989690722	0.945812808
62	0.906542056	1	0.950980392
63	0.906542056	1	0.950980392
64	0.906542056	1	0.950980392
65	0.906542056	1	0.950980392
66	0.906542056	1	0.950980392
67	0.906542056	1	0.950980392
68	0.906542056	1	0.950980392
69	0.906542056	1	0.950980392
70	0.906542056	1	0.950980392
71	0.906542056	1	0.950980392
72	0.906542056	1	0.950980392
73	0.906542056	1	0.950980392
74	0.906542056	1	0.950980392
75	0.906542056	1	0.950980392
76	0.906542056	1	0.950980392
77	0.906542056	1	0.950980392
78	0.906542056	1	0.950980392
79	0.906542056	1	0.950980392
80	0.906542056	1	0.950980392
81	0.906542056	1	0.950980392
82	0.906542056	1	0.950980392
83	0.906542056	1	0.950980392
84	0.906542056	1	0.950980392

Table G-38. P, R, F-Score for 20s

8. Extracted Test Data: 30s and 40s

20-29	Precision	Recall	F-Score
Baseline	0.906542056	1	0.950980392
1	0.906542056	1	0.950980392
2	0.906542056	1	0.950980392
3	0.906542056	1	0.950980392
4	0.906542056	1	0.950980392
5	0.906542056	1	0.950980392
6	0.906542056	1	0.950980392
7	0.906542056	1	0.950980392
8	0.906542056	1	0.950980392
9	0.906542056	1	0.950980392
10	0.906542056	1	0.950980392
11	0.906542056	1	0.950980392
12	0.906542056	1	0.950980392
13	0.906542056	1	0.950980392
14	0.905660377	0.989690722	0.945812808
15	0.905660377	0.989690722	0.945812808
16	0.906542056	1	0.950980392
17	0.906542056	1	0.950980392
18	0.906542056	1	0.950980392
19	0.906542056	1	0.950980392
20	0.906542056	1	0.950980392
21	0.906542056	1	0.950980392
22	0.905660377	0.989690722	0.945812808
23	0.906542056	1	0.950980392
24	0.906542056	1	0.950980392
25	0.906542056	1	0.950980392
26	0.906542056	1	0.950980392
27	0.905660377	0.989690722	0.945812808
28	0.904761905	0.979381443	0.940594059
29	0.906542056	1	0.950980392
30	0.906542056	1	0.950980392
31	0.906542056	1	0.950980392
32	0.906542056	1	0.950980392
33	0.906542056	1	0.950980392
34	0.905660377	0.989690722	0.945812808
35	0.906542056	1	0.950980392
36	0.906542056	1	0.950980392
37	0.906542056	1	0.950980392
38	0.906542056	1	0.950980392
39	0.906542056	1	0.950980392
40	0.906542056	1	0.950980392
41	0.906542056	1	0.950980392

42	0.906542056	1	0.950980392
43	0.906542056	1	0.950980392
44	0.906542056	1	0.950980392
45	0.906542056	1	0.950980392
46	0.906542056	1	0.950980392
47	0.906542056	1	0.950980392
48	0.906542056	1	0.950980392
49	0.906542056	1	0.950980392
50	0.906542056	1	0.950980392
51	0.906542056	1	0.950980392
52	0.906542056	1	0.950980392
53	0.906542056	1	0.950980392
54	0.906542056	1	0.950980392
55	0.906542056	1	0.950980392
56	0.906542056	1	0.950980392
57	0.906542056	1	0.950980392
58	0.906542056	1	0.950980392
59	0.906542056	1	0.950980392
60	0.906542056	1	0.950980392
61	0.905660377	0.989690722	0.945812808
62	0.906542056	1	0.950980392
63	0.906542056	1	0.950980392
64	0.906542056	1	0.950980392
65	0.906542056	1	0.950980392
66	0.906542056	1	0.950980392
67	0.906542056	1	0.950980392
68	0.906542056	1	0.950980392
69	0.906542056	1	0.950980392
70	0.906542056	1	0.950980392
71	0.906542056	1	0.950980392
72	0.906542056	1	0.950980392
73	0.906542056	1	0.950980392
74	0.906542056	1	0.950980392
75	0.906542056	1	0.950980392
76	0.906542056	1	0.950980392
77	0.906542056	1	0.950980392
78	0.906542056	1	0.950980392
79	0.906542056	1	0.950980392
80	0.906542056	1	0.950980392
81	0.906542056	1	0.950980392
82	0.906542056	1	0.950980392
83	0.906542056	1	0.950980392
84	0.906542056	1	0.950980392

Table G-39. P, R, F-Score for 30s

40-49	Precision	Recall	F-Score
Baseline	0.463768116	1	0.633663366
2	1	0.0625	0.117647059
4	0.5	0.03125	0.058823529
15	1	0.03125	0.060606061
22	1	0.03125	0.060606061
59	1	0.03125	0.060606061
68	1	0.03125	0.060606061
78	0.677419355	0.65625	0.666666667

Table G-40. P, R, F-Score for 40s

9. Extracted Test Data: 30s and 50s

30-39	Precision	Recall	F-Score
Baseline	0.787234043	1	0.880952381
1	0.782608696	0.972972973	0.86746988
2	0.787234043	1	0.880952381
3	0.787234043	1	0.880952381
4	0.787234043	1	0.880952381
5	0.787234043	1	0.880952381
6	0.787234043	1	0.880952381
7	0.787234043	1	0.880952381
8	0.787234043	1	0.880952381
9	0.787234043	1	0.880952381
10	0.782608696	0.972972973	0.86746988
11	0.787234043	1	0.880952381
12	0.787234043	1	0.880952381
13	0.787234043	1	0.880952381
14	0.782608696	0.972972973	0.86746988
15	0.787234043	1	0.880952381
16	0.787234043	1	0.880952381
17	0.787234043	1	0.880952381
18	0.787234043	1	0.880952381
19	0.787234043	1	0.880952381
20	0.787234043	1	0.880952381
21	0.787234043	1	0.880952381
22	0.787234043	1	0.880952381
23	0.787234043	1	0.880952381
24	0.787234043	1	0.880952381
25	0.787234043	1	0.880952381
26	0.787234043	1	0.880952381
27	0.787234043	1	0.880952381
28	0.787234043	1	0.880952381
29	0.787234043	1	0.880952381
30	0.787234043	1	0.880952381
31	0.787234043	1	0.880952381
32	0.782608696	0.972972973	0.86746988
33	0.787234043	1	0.880952381
34	0.787234043	1	0.880952381
35	0.782608696	0.972972973	0.86746988
36	0.787234043	1	0.880952381
37	0.782608696	0.972972973	0.86746988
38	0.787234043	1	0.880952381
39	0.787234043	1	0.880952381
40	0.787234043	1	0.880952381
41	0.787234043	1	0.880952381
42	0.787234043	1	0.880952381

43	0.787234043	1	0.880952381
44	0.787234043	1	0.880952381
45	0.787234043	1	0.880952381
46	0.787234043	1	0.880952381
47	0.782608696	0.972972973	0.86746988
48	0.787234043	1	0.880952381
49	0.782608696	0.972972973	0.86746988
50	0.782608696	0.972972973	0.86746988
51	0.787234043	1	0.880952381
52	0.787234043	1	0.880952381
53	0.787234043	1	0.880952381
54	0.787234043	1	0.880952381
55	0.787234043	1	0.880952381
56	0.782608696	0.972972973	0.86746988
57	0.787234043	1	0.880952381
58	0.787234043	1	0.880952381
59	0.787234043	1	0.880952381
60	0.787234043	1	0.880952381
61	0.787234043	1	0.880952381
62	0.787234043	1	0.880952381
63	0.787234043	1	0.880952381
64	0.787234043	1	0.880952381
65	0.787234043	1	0.880952381
66	0.787234043	1	0.880952381
67	0.787234043	1	0.880952381
68	0.787234043	1	0.880952381
69	0.787234043	1	0.880952381
70	0.787234043	1	0.880952381
71	0.787234043	1	0.880952381
72	0.787234043	1	0.880952381
73	0.787234043	1	0.880952381
74	0.787234043	1	0.880952381
75	0.787234043	1	0.880952381
76	0.787234043	1	0.880952381
77	0.787234043	1	0.880952381
78	0.787234043	1	0.880952381
79	0.787234043	1	0.880952381
80	0.787234043	1	0.880952381
81	0.787234043	1	0.880952381
82	0.787234043	1	0.880952381
83	0.787234043	1	0.880952381
84	0.787234043	1	0.880952381

Table G-41. P, R, F-Score for 30s

10. Extracted Test Data: 40s and 50s

40-49	Precision	Recall	F-Score
Baseline	0.761904762	1	0.864864865
1	0.761904762	1	0.864864865
2	0.761904762	1	0.864864865
3	0.761904762	1	0.864864865
4	0.761904762	1	0.864864865
5	0.761904762	1	0.864864865
6	0.756097561	0.96875	0.849315068
7	0.761904762	1	0.864864865
8	0.761904762	1	0.864864865
9	0.761904762	1	0.864864865
10	0.761904762	1	0.864864865
11	0.761904762	1	0.864864865
12	0.761904762	1	0.864864865
13	0.761904762	1	0.864864865
14	0.761904762	1	0.864864865
15	0.761904762	1	0.864864865
16	0.761904762	1	0.864864865
17	0.761904762	1	0.864864865
18	0.780487805	1	0.876712329
19	0.761904762	1	0.864864865
20	0.761904762	1	0.864864865
21	0.761904762	1	0.864864865
22	0.761904762	1	0.864864865
23	0.761904762	1	0.864864865
24	0.761904762	1	0.864864865
25	0.743589744	0.90625	0.816901408
26	0.743589744	0.90625	0.816901408
27	0.780487805	1	0.876712329
28	0.761904762	1	0.864864865
29	0.761904762	1	0.864864865
30	0.761904762	1	0.864864865
31	0.761904762	1	0.864864865
32	0.761904762	1	0.864864865
33	0.761904762	1	0.864864865
34	0.761904762	1	0.864864865
35	0.761904762	1	0.864864865
36	0.761904762	1	0.864864865
37	0.761904762	1	0.864864865
38	0.761904762	1	0.864864865
39	0.756097561	0.96875	0.849315068
40	0.761904762	1	0.864864865
41	0.761904762	1	0.864864865
42	0.761904762	1	0.864864865

43	0.761904762	1	0.864864865
44	0.761904762	1	0.864864865
45	0.756097561	0.96875	0.849315068
46	0.756097561	0.96875	0.849315068
47	0.761904762	1	0.864864865
48	0.761904762	1	0.864864865
49	0.761904762	1	0.864864865
50	0.761904762	1	0.864864865
51	0.761904762	1	0.864864865
52	0.761904762	1	0.864864865
53	0.761904762	1	0.864864865
54	0.761904762	1	0.864864865
55	0.761904762	1	0.864864865
56	0.761904762	1	0.864864865
57	0.761904762	1	0.864864865
58	0.761904762	1	0.864864865
59	0.761904762	1	0.864864865
60	0.756097561	0.96875	0.849315068
61	0.761904762	1	0.864864865
62	0.761904762	1	0.864864865
63	0.761904762	1	0.864864865
64	0.761904762	1	0.864864865
65	0.761904762	1	0.864864865
66	0.761904762	1	0.864864865
67	0.761904762	1	0.864864865
68	0.761904762	1	0.864864865
69	0.761904762	1	0.864864865
70	0.761904762	1	0.864864865
71	0.761904762	1	0.864864865
72	0.761904762	1	0.864864865
73	0.761904762	1	0.864864865
74	0.761904762	1	0.864864865
75	0.761904762	1	0.864864865
76	0.761904762	1	0.864864865
77	0.761904762	1	0.864864865
78	0.761904762	1	0.864864865
79	0.761904762	1	0.864864865
80	0.761904762	1	0.864864865
81	0.761904762	1	0.864864865
82	0.761904762	1	0.864864865
83	0.761904762	1	0.864864865
84	0.761904762	1	0.864864865

Table G-42. P, R, F-Score for 40s

50-59	Precision	Recall	F-Score
Baseline	0.238095238	1	0.384615385
18	1	0.1	0.181818182
27	1	0.1	0.181818182

Table G-43. P, R, F-Score for 40s

11. Extracted Test Data: Under 26 and 26 or Over

< 26	Precision	Recall	F-Score
Baseline	0.540983607	1	0.70212766
1	0.540983607	1	0.70212766
2	0.540983607	1	0.70212766
3	0.543209877	1	0.704
4	0.543209877	1	0.704
5	0.540983607	1	0.70212766
6	0.540983607	1	0.70212766
7	0.540983607	1	0.70212766
8	0.540983607	1	0.70212766
9	0.540983607	1	0.70212766
10	0.536170213	0.954545455	0.686648501
11	0.53909465	0.992424242	0.698666667
12	0.548117155	0.992424242	0.706199461
13	0.540983607	1	0.70212766
14	0.541322314	0.992424242	0.700534759
15	0.540983607	1	0.70212766
16	0.540983607	1	0.70212766
17	0.540983607	1	0.70212766
18	0.540983607	1	0.70212766
19	0.540983607	1	0.70212766
20	0.540983607	1	0.70212766
21	0.540983607	1	0.70212766
22	0.540983607	1	0.70212766
23	0.540983607	1	0.70212766
24	0.537190083	0.984848485	0.695187166
25	0.53909465	0.992424242	0.698666667
26	0.540983607	1	0.70212766
27	0.53909465	0.992424242	0.698666667
28	0.53909465	0.992424242	0.698666667
29	0.540983607	1	0.70212766
30	0.540983607	1	0.70212766
31	0.53909465	0.992424242	0.698666667
32	0.53909465	0.992424242	0.698666667
33	0.53909465	0.992424242	0.698666667
34	0.537190083	0.984848485	0.695187166
35	0.540983607	1	0.70212766
36	0.540983607	1	0.70212766
37	0.53909465	0.992424242	0.698666667
38	0.541322314	0.992424242	0.700534759
39	0.541322314	0.992424242	0.700534759
40	0.540983607	1	0.70212766
41	0.540983607	1	0.70212766
42	0.540983607	1	0.70212766

43	0.540983607	1	0.70212766
44	0.540983607	1	0.70212766
45	0.540983607	1	0.70212766
46	0.540983607	1	0.70212766
47	0.540983607	1	0.70212766
48	0.540983607	1	0.70212766
49	0.540983607	1	0.70212766
50	0.543209877	1	0.704
51	0.540983607	1	0.70212766
52	0.540983607	1	0.70212766
53	0.543209877	1	0.704
54	0.540983607	1	0.70212766
55	0.53909465	0.992424242	0.698666667
56	0.531380753	0.962121212	0.684636119
57	0.540983607	1	0.70212766
58	0.540983607	1	0.70212766
59	0.540983607	1	0.70212766
60	0.540983607	1	0.70212766
61	0.540983607	1	0.70212766
62	0.540983607	1	0.70212766
63	0.540983607	1	0.70212766
64	0.540983607	1	0.70212766
65	0.540983607	1	0.70212766
66	0.540983607	1	0.70212766
67	0.540983607	1	0.70212766
68	0.540983607	1	0.70212766
69	0.540983607	1	0.70212766
70	0.540983607	1	0.70212766
71	0.540983607	1	0.70212766
72	0.540983607	1	0.70212766
73	0.540983607	1	0.70212766
74	0.540983607	1	0.70212766
75	0.540983607	1	0.70212766
76	0.540983607	1	0.70212766
77	0.540983607	1	0.70212766
78	0.540983607	1	0.70212766
79	0.540983607	1	0.70212766
80	0.540983607	1	0.70212766
81	0.540983607	1	0.70212766
82	0.540983607	1	0.70212766
83	0.540983607	1	0.70212766
84	0.540983607	1	0.70212766

Table G-44. P, R, F-Score for Under 26

>= 26	Precision	Recall	F-Score
Baseline	0.459016393	1	0.629213483
3	1	0.008928571	0.017699115
4	1	0.008928571	0.017699115
10	0.333333333	0.026785714	0.049586777
12	0.8	0.035714286	0.068376068
14	0.5	0.008928571	0.01754386
38	0.5	0.008928571	0.01754386
39	0.5	0.008928571	0.01754386
50	1	0.008928571	0.017699115
53	1	0.008928571	0.017699115

Table G-45. P, R, F-Score for 26 and Older

D. GENDER: BINARY CLASSIFICATION WITHOUT PRIOR

1. All Test Data

Male	Precision	Recall	F-Score
Feature	0.525835866	1	0.689243028
2	0.53	0.919075145	0.67230444
3	0.524390244	0.994219653	0.686626747
4	0.524390244	0.994219653	0.686626747
5	0.525835866	1	0.689243028
6	0.525835866	1	0.689243028
7	0.525835866	1	0.689243028
8	0.525835866	1	0.689243028
9	0.526315789	0.98265896	0.685483871
10	0.52866242	0.959537572	0.681724846
11	0.526153846	0.988439306	0.686746988
12	0.357142857	0.028901734	0.053475936
13	0.525835866	1	0.689243028
14	0.524390244	0.994219653	0.686626747
15	0.5	0.052023121	0.094240838
16	0.642857143	0.052023121	0.096256684
17	0.5	0.005780347	0.011428571
18	0.5	0.005780347	0.011428571
19	0.525835866	1	0.689243028
21	0.525835866	1	0.689243028
22	0.525993884	0.994219653	0.688
23	0.525835866	1	0.689243028
24	0.525835866	1	0.689243028
25	0.529051988	1	0.692
26	0.529051988	1	0.692
27	0.527439024	1	0.690618762
28	0.526153846	0.988439306	0.686746988
29	0.527439024	1	0.690618762
30	0.525835866	1	0.689243028
31	0.524390244	0.994219653	0.686626747
32	0.524390244	0.994219653	0.686626747
33	0.527439024	1	0.690618762
34	0.525835866	1	0.689243028
35	0.52293578	0.988439306	0.684
37	0.525835866	1	0.689243028
38	0.529051988	1	0.692
39	0.525835866	1	0.689243028
41	0.524390244	0.994219653	0.686626747
43	0.525835866	1	0.689243028
44	0.525835866	1	0.689243028
45	0.525835866	1	0.689243028
46	0.525835866	1	0.689243028

47	0.527439024	1	0.690618762
49	0.525835866	1	0.689243028
50	0.525835866	1	0.689243028
51	0.525835866	1	0.689243028
52	0.525835866	1	0.689243028
53	1	0.005780347	0.011494253
55	0.525835866	1	0.689243028
56	0.529595016	0.98265896	0.688259109
58	0.525835866	1	0.689243028
59	0.525835866	1	0.689243028
62	0.516129032	0.092485549	0.156862745
63	0.525835866	1	0.689243028
64	0.538461538	0.040462428	0.075268817
65	0.525835866	1	0.689243028
66	0.525835866	1	0.689243028
67	0.525835866	1	0.689243028
68	0.524390244	0.994219653	0.686626747
69	0.527439024	1	0.690618762
71	0.525835866	1	0.689243028
72	0.525835866	1	0.689243028
73	0.52293578	0.988439306	0.684
76	0.525835866	1	0.689243028
77	0.444444444	0.069364162	0.12
78	0.556701031	0.312138728	0.4
79	0.527439024	1	0.690618762
80	0.525835866	1	0.689243028
81	0.527439024	1	0.690618762
82	0.525835866	1	0.689243028
83	0.525835866	1	0.689243028

Table G-46. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.474164134	1	0.643298969
1	0.472560976	0.993589744	0.640495868
2	0.517241379	0.096153846	0.162162162
9	0.5	0.019230769	0.037037037
10	0.533333333	0.051282051	0.093567251
11	0.5	0.012820513	0.025
12	0.466666667	0.942307692	0.624203822
15	0.47266881	0.942307692	0.629550321
16	0.479365079	0.967948718	0.64118896
17	0.474006116	0.993589744	0.641821946
18	0.474006116	0.993589744	0.641821946
20	0.470948012	0.987179487	0.637681159
22	0.5	0.006410256	0.012658228
25	1	0.012820513	0.025316456
26	1	0.012820513	0.025316456
27	1	0.006410256	0.012738854
28	0.5	0.012820513	0.025
29	1	0.006410256	0.012738854
33	1	0.006410256	0.012738854
36	0.474164134	1	0.643298969
38	1	0.012820513	0.025316456
40	0.474164134	1	0.643298969
42	0.474164134	1	0.643298969
47	1	0.006410256	0.012738854
48	0.472560976	0.993589744	0.640495868
53	0.475609756	1	0.644628099
54	0.474164134	1	0.643298969
56	0.625	0.032051282	0.06097561
57	0.474164134	1	0.643298969
60	0.474164134	1	0.643298969
61	0.474164134	1	0.643298969
62	0.473154362	0.903846154	0.621145374
64	0.474683544	0.961538462	0.63559322
69	1	0.006410256	0.012738854
70	0.474164134	1	0.643298969
74	0.474164134	1	0.643298969
75	0.474164134	1	0.643298969
77	0.466887417	0.903846154	0.615720524
78	0.487068966	0.724358974	0.582474227
79	1	0.006410256	0.012738854
81	1	0.006410256	0.012738854
84	0.474164134	1	0.643298969

Table G-47.

P, R, F-Score for Females

2. Extracted Test Data: Teens and 20s

Feature	Precision	Recall	F-Score
Male	0.448484848	1	0.619246862
2	0.46	0.932432432	0.616071429
3	0.448484848	1	0.619246862
4	0.448484848	1	0.619246862
8	0.448484848	1	0.619246862
9	0.445859873	0.945945946	0.606060606
10	0.428571429	0.040540541	0.074074074
11	0.45398773	1	0.624472574
12	0.2	0.013513514	0.025316456
14	0.450617284	0.986486486	0.618644068
15	0.230769231	0.040540541	0.068965517
16	0.545454545	0.081081081	0.141176471
17	0.450617284	0.986486486	0.618644068
18	0.456790123	1	0.627118644
22	0.445121951	0.986486486	0.613445378
24	0.5	0.013513514	0.026315789
27	0.441717791	0.972972973	0.607594937
28	1	0.013513514	0.026666667
31	0.448484848	1	0.619246862
32	0.448484848	1	0.619246862
33	0.5	0.013513514	0.026315789
34	0.333333333	0.013513514	0.025974026
39	0.445121951	0.986486486	0.613445378
42	0.448484848	1	0.619246862
43	0.448484848	1	0.619246862
45	0.448484848	1	0.619246862
51	0.448484848	1	0.619246862
52	0.448484848	1	0.619246862
53	1	0.013513514	0.026666667
55	0.5	0.013513514	0.026315789
56	0.4625	1	0.632478632
57	0.5	0.013513514	0.026315789
62	0.441558442	0.918918919	0.596491228
64	0.416666667	0.067567568	0.11627907
68	0.448484848	1	0.619246862
71	1	0.013513514	0.026666667
73	1	0.013513514	0.026666667
77	0.326923077	0.22972973	0.26984127
78	0.493506494	0.513513514	0.503311258

Table G-48. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.551515152	1	0.7109375
1	0.551515152	1	0.7109375
2	0.666666667	0.10989011	0.188679245
5	0.551515152	1	0.7109375
6	0.551515152	1	0.7109375
7	0.551515152	1	0.7109375
9	0.5	0.043956044	0.080808081
10	0.550632911	0.956043956	0.698795181
11	1	0.021978022	0.043010753
12	0.54375	0.956043956	0.693227092
13	0.551515152	1	0.7109375
14	0.666666667	0.021978022	0.042553191
15	0.532894737	0.89010989	0.666666667
16	0.558441558	0.945054945	0.702040816
17	0.666666667	0.021978022	0.042553191
18	1	0.032967033	0.063829787
19	0.551515152	1	0.7109375
20	0.551515152	1	0.7109375
21	0.551515152	1	0.7109375
23	0.548780488	0.989010989	0.705882353
24	0.552147239	0.989010989	0.708661417
25	0.548780488	0.989010989	0.705882353
26	0.548780488	0.989010989	0.705882353
28	0.554878049	1	0.71372549
29	0.548780488	0.989010989	0.705882353
30	0.551515152	1	0.7109375
33	0.552147239	0.989010989	0.708661417
34	0.549382716	0.978021978	0.703557312
35	0.548780488	0.989010989	0.705882353
36	0.551515152	1	0.7109375
37	0.548780488	0.989010989	0.705882353
38	0.548780488	0.989010989	0.705882353
40	0.551515152	1	0.7109375
41	0.551515152	1	0.7109375
44	0.551515152	1	0.7109375
46	0.551515152	1	0.7109375
47	0.551515152	1	0.7109375
48	0.548780488	0.989010989	0.705882353
49	0.551515152	1	0.7109375
50	0.551515152	1	0.7109375
53	0.554878049	1	0.71372549
54	0.551515152	1	0.7109375
55	0.552147239	0.989010989	0.708661417
56	1	0.054945055	0.104166667

57	0.552147239	0.989010989	0.708661417
58	0.551515152	1	0.7109375
59	0.551515152	1	0.7109375
60	0.551515152	1	0.7109375
61	0.551515152	1	0.7109375
62	0.454545455	0.054945055	0.098039216
63	0.548780488	0.989010989	0.705882353
64	0.549019608	0.923076923	0.68852459
65	0.551515152	1	0.7109375
66	0.551515152	1	0.7109375
67	0.551515152	1	0.7109375
69	0.548780488	0.989010989	0.705882353
70	0.548780488	0.989010989	0.705882353
71	0.554878049	1	0.71372549
72	0.551515152	1	0.7109375
73	0.554878049	1	0.71372549
74	0.551515152	1	0.7109375
75	0.551515152	1	0.7109375
76	0.551515152	1	0.7109375
77	0.495575221	0.615384615	0.549019608
78	0.590909091	0.571428571	0.581005587
79	0.551515152	1	0.7109375
80	0.551515152	1	0.7109375
81	0.551515152	1	0.7109375
82	0.551515152	1	0.7109375
83	0.551515152	1	0.7109375
84	0.551515152	1	0.7109375

Table G-49. P, R, F-Score for Females

3. Extracted Test Data: Teens and 30s

Male	Precision	Recall	F-Score
Baseline	0.428571429	1	0.6
1	1	0.022222222	0.043478261
2	0.445544554	1	0.616438356
3	1	0.022222222	0.043478261
4	1	0.022222222	0.043478261
7	0.428571429	1	0.6
9	0.5	0.022222222	0.042553191
10	0.436893204	1	0.608108108
11	0.5	0.044444444	0.081632653
12	0.25	0.022222222	0.040816327
13	0.427184466	0.977777778	0.594594595
14	0.428571429	1	0.6
15	0.5	0.044444444	0.081632653
16	0.8	0.088888889	0.16
17	0.436893204	1	0.608108108
18	0.441176471	1	0.612244898
31	0.428571429	1	0.6
32	0.428571429	1	0.6
34	0.436893204	1	0.608108108
35	0.436893204	1	0.608108108
37	0.5	0.022222222	0.042553191
38	0.5	0.022222222	0.042553191
42	0.428571429	1	0.6
46	0.432692308	1	0.604026846
47	0.432692308	1	0.604026846
51	0.428571429	1	0.6
55	0.435643564	0.977777778	0.602739726
56	0.435643564	0.977777778	0.602739726
62	0.422680412	0.911111111	0.577464789
64	0.333333333	0.044444444	0.078431373
71	0.428571429	1	0.6
77	0.266666667	0.177777778	0.213333333
78	0.35	0.311111111	0.329411765

Table G-50.

P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.571428571	1	0.727272727
1	0.576923077	1	0.731707317
2	1	0.066666667	0.125
3	0.576923077	1	0.731707317
4	0.576923077	1	0.731707317
5	0.571428571	1	0.727272727
6	0.571428571	1	0.727272727
8	0.571428571	1	0.727272727
9	0.572815534	0.983333333	0.72392638
10	1	0.033333333	0.064516129
11	0.574257426	0.966666667	0.720496894
12	0.564356436	0.95	0.708074534
13	0.5	0.016666667	0.032258065
15	0.574257426	0.966666667	0.720496894
16	0.59	0.983333333	0.7375
17	1	0.033333333	0.064516129
18	1	0.05	0.095238095
19	0.571428571	1	0.727272727
20	0.567307692	0.983333333	0.719512195
21	0.571428571	1	0.727272727
22	0.571428571	1	0.727272727
23	0.571428571	1	0.727272727
24	0.571428571	1	0.727272727
25	0.567307692	0.983333333	0.719512195
26	0.571428571	1	0.727272727
27	0.571428571	1	0.727272727
28	0.571428571	1	0.727272727
29	0.567307692	0.983333333	0.719512195
30	0.567307692	0.983333333	0.719512195
33	0.567307692	0.983333333	0.719512195
34	1	0.033333333	0.064516129
35	1	0.033333333	0.064516129
36	0.571428571	1	0.727272727
37	0.572815534	0.983333333	0.72392638
38	0.572815534	0.983333333	0.72392638
39	0.571428571	1	0.727272727
40	0.571428571	1	0.727272727
41	0.571428571	1	0.727272727
43	0.571428571	1	0.727272727
44	0.571428571	1	0.727272727
45	0.571428571	1	0.727272727
46	1	0.016666667	0.032786885
47	1	0.016666667	0.032786885

48	0.567307692	0.9833333333	0.719512195
49	0.571428571	1	0.727272727
50	0.571428571	1	0.727272727
52	0.571428571	1	0.727272727
53	0.563106796	0.966666667	0.711656442
54	0.571428571	1	0.727272727
55	0.75	0.05	0.09375
56	0.75	0.05	0.09375
57	0.567307692	0.9833333333	0.719512195
58	0.571428571	1	0.727272727
59	0.571428571	1	0.727272727
60	0.571428571	1	0.727272727
61	0.571428571	1	0.727272727
62	0.5	0.066666667	0.117647059
63	0.567307692	0.9833333333	0.719512195
64	0.565656566	0.9333333333	0.704402516
65	0.571428571	1	0.727272727
66	0.571428571	1	0.727272727
67	0.571428571	1	0.727272727
68	0.571428571	1	0.727272727
69	0.567307692	0.9833333333	0.719512195
70	0.567307692	0.9833333333	0.719512195
72	0.571428571	1	0.727272727
73	0.571428571	1	0.727272727
74	0.571428571	1	0.727272727
75	0.571428571	1	0.727272727
76	0.571428571	1	0.727272727
77	0.506666667	0.6333333333	0.562962963
78	0.523076923	0.566666667	0.544
79	0.571428571	1	0.727272727
80	0.571428571	1	0.727272727
81	0.571428571	1	0.727272727
82	0.571428571	1	0.727272727
83	0.571428571	1	0.727272727
84	0.571428571	1	0.727272727

Table G-51. P, R, F-Score for Females

4. Extracted Test Data: Teens and 40s

Male	Precision	Recall	F-Score
Baseline	0.42	1	0.591549296
2	0.430107527	0.952380952	0.592592593
3	0.424242424	1	0.595744681
4	0.424242424	1	0.595744681
9	0.416666667	0.952380952	0.579710145
10	0.418367347	0.976190476	0.585714286
11	1	0.023809524	0.046511628
14	0.424242424	1	0.595744681
15	0.411111111	0.880952381	0.560606061
16	0.431578947	0.976190476	0.598540146
17	0.428571429	1	0.6
18	0.432989691	1	0.604316547
26	0.5	0.023809524	0.045454545
27	0.42	1	0.591549296
31	0.42	1	0.591549296
32	0.42	1	0.591549296
34	0.428571429	1	0.6
35	0.428571429	1	0.6
39	1	0.023809524	0.046511628
42	1	0.023809524	0.046511628
55	0.424242424	1	0.595744681
56	0.432989691	1	0.604316547
59	1	0.023809524	0.046511628
62	0.5	0.095238095	0.16
64	0.333333333	0.023809524	0.044444444
67	0.42	1	0.591549296
68	1	0.023809524	0.046511628
71	1	0.023809524	0.046511628
77	0.379310345	0.261904762	0.309859155
78	0.375	0.285714286	0.324324324

Table G-52.

P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.58	1	0.734177215
1	0.58	1	0.734177215
2	0.714285714	0.086206897	0.153846154
3	1	0.017241379	0.033898305
4	1	0.017241379	0.033898305
5	0.58	1	0.734177215
6	0.58	1	0.734177215
7	0.58	1	0.734177215
8	0.58	1	0.734177215
9	0.5	0.034482759	0.064516129
10	0.5	0.017241379	0.033333333
11	0.585858586	1	0.738853503
12	0.575757576	0.982758621	0.72611465
13	0.58	1	0.734177215
14	1	0.017241379	0.033898305
15	0.5	0.086206897	0.147058824
16	0.8	0.068965517	0.126984127
17	1	0.034482759	0.066666667
18	1	0.051724138	0.098360656
19	0.58	1	0.734177215
20	0.58	1	0.734177215
21	0.58	1	0.734177215
22	0.58	1	0.734177215
23	0.58	1	0.734177215
24	0.58	1	0.734177215
25	0.575757576	0.982758621	0.72611465
26	0.581632653	0.982758621	0.730769231
28	0.58	1	0.734177215
29	0.575757576	0.982758621	0.72611465
30	0.575757576	0.982758621	0.72611465
33	0.575757576	0.982758621	0.72611465
34	1	0.034482759	0.066666667
35	1	0.034482759	0.066666667
36	0.58	1	0.734177215
37	0.575757576	0.982758621	0.72611465
38	0.575757576	0.982758621	0.72611465
39	0.585858586	1	0.738853503
40	0.58	1	0.734177215
41	0.58	1	0.734177215
42	0.585858586	1	0.738853503
43	0.58	1	0.734177215
44	0.58	1	0.734177215
45	0.58	1	0.734177215

46	0.575757576	0.982758621	0.72611465
47	0.58	1	0.734177215
48	0.575757576	0.982758621	0.72611465
49	0.58	1	0.734177215
50	0.58	1	0.734177215
51	0.58	1	0.734177215
52	0.58	1	0.734177215
53	0.58	1	0.734177215
54	0.58	1	0.734177215
55	1	0.017241379	0.033898305
56	1	0.051724138	0.098360656
57	0.575757576	0.982758621	0.72611465
58	0.58	1	0.734177215
59	0.585858586	1	0.738853503
60	0.58	1	0.734177215
61	0.58	1	0.734177215
62	0.586956522	0.931034483	0.72
63	0.575757576	0.982758621	0.72611465
64	0.577319588	0.965517241	0.722580645
65	0.58	1	0.734177215
66	0.58	1	0.734177215
68	0.585858586	1	0.738853503
69	0.575757576	0.982758621	0.72611465
70	0.575757576	0.982758621	0.72611465
71	0.585858586	1	0.738853503
72	0.58	1	0.734177215
73	0.58	1	0.734177215
74	0.58	1	0.734177215
75	0.58	1	0.734177215
76	0.58	1	0.734177215
77	0.563380282	0.689655172	0.620155039
78	0.558823529	0.655172414	0.603174603
79	0.58	1	0.734177215
80	0.58	1	0.734177215
81	0.58	1	0.734177215
82	0.58	1	0.734177215
83	0.58	1	0.734177215
84	0.58	1	0.734177215

Table G-53. P, R, F-Score for Females

5. Extracted Test Data: Teens and 50s

Male	Precision	Recall	F-Score
Baseline	0.384615385	1	0.555555556
2	0.416666667	1	0.588235294
3	0.384615385	1	0.555555556
4	0.384615385	1	0.555555556
9	0.394736842	1	0.566037736
10	0.38961039	1	0.560747664
11	0.75	0.1	0.176470588
12	0.666666667	0.066666667	0.121212121
14	0.38961039	1	0.560747664
15	0	0	0
16	0.397260274	0.966666667	0.563106796
18	0.384615385	1	0.555555556
31	0.384615385	1	0.555555556
32	0.384615385	1	0.555555556
34	0.394736842	1	0.566037736
42	1	0.033333333	0.064516129
52	1	0.033333333	0.064516129
56	0.4	1	0.571428571
57	0.5	0.033333333	0.0625
62	0.386666667	0.966666667	0.552380952
71	1	0.033333333	0.064516129
77	0.230769231	0.2	0.214285714
78	0.318181818	0.233333333	0.269230769

Table G-54.

P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.615384615	1	0.761904762
1	0.6	0.9375	0.731707317
2	1	0.125	0.222222222
5	0.615384615	1	0.761904762
6	0.615384615	1	0.761904762
7	0.615384615	1	0.761904762
8	0.615384615	1	0.761904762
9	1	0.041666667	0.08
10	1	0.020833333	0.040816327
11	0.635135135	0.979166667	0.770491803
12	0.626666667	0.979166667	0.764227642
13	0.615384615	1	0.761904762
14	1	0.020833333	0.040816327
15	0.615384615	1	0.761904762
16	0.8	0.083333333	0.150943396
17	0.615384615	1	0.761904762
19	0.615384615	1	0.761904762
20	0.615384615	1	0.761904762
21	0.615384615	1	0.761904762
22	0.615384615	1	0.761904762
23	0.61038961	0.979166667	0.752
24	0.61038961	0.979166667	0.752
25	0.61038961	0.979166667	0.752
26	0.61038961	0.979166667	0.752
27	0.615384615	1	0.761904762
28	0.615384615	1	0.761904762
29	0.61038961	0.979166667	0.752
30	0.61038961	0.979166667	0.752
33	0.61038961	0.979166667	0.752
34	1	0.041666667	0.08
35	0.61038961	0.979166667	0.752
36	0.615384615	1	0.761904762
37	0.61038961	0.979166667	0.752
38	0.61038961	0.979166667	0.752
39	0.615384615	1	0.761904762
40	0.615384615	1	0.761904762
41	0.615384615	1	0.761904762
42	0.623376623	1	0.768
43	0.615384615	1	0.761904762
44	0.615384615	1	0.761904762
45	0.615384615	1	0.761904762
46	0.61038961	0.979166667	0.752
47	0.615384615	1	0.761904762

48	0.61038961	0.979166667	0.752
49	0.615384615	1	0.761904762
50	0.615384615	1	0.761904762
51	0.615384615	1	0.761904762
52	0.623376623	1	0.768
53	0.615384615	1	0.761904762
54	0.615384615	1	0.761904762
55	0.61038961	0.979166667	0.752
56	1	0.0625	0.117647059
57	0.618421053	0.979166667	0.758064516
58	0.615384615	1	0.761904762
59	0.615384615	1	0.761904762
60	0.615384615	1	0.761904762
61	0.615384615	1	0.761904762
62	0.666666667	0.041666667	0.078431373
63	0.61038961	0.979166667	0.752
64	0.615384615	1	0.761904762
65	0.615384615	1	0.761904762
66	0.615384615	1	0.761904762
67	0.615384615	1	0.761904762
68	0.615384615	1	0.761904762
69	0.61038961	0.979166667	0.752
70	0.61038961	0.979166667	0.752
71	0.623376623	1	0.768
72	0.615384615	1	0.761904762
73	0.615384615	1	0.761904762
74	0.615384615	1	0.761904762
75	0.615384615	1	0.761904762
76	0.615384615	1	0.761904762
77	0.538461538	0.583333333	0.56
78	0.589285714	0.6875	0.634615385
79	0.615384615	1	0.761904762
80	0.615384615	1	0.761904762
81	0.615384615	1	0.761904762
82	0.615384615	1	0.761904762
83	0.615384615	1	0.761904762
84	0.615384615	1	0.761904762

Table G-55. P, R, F-Score for Females

6. Extracted Test Data: 20s and 30s

Male	Precision	Recall	F-Score
Baseline	0.559701493	1	0.717703349
1	0.557251908	0.973333333	0.708737864
2	0.56	0.933333333	0.7
3	0.556390977	0.986666667	0.711538462
4	0.556390977	0.986666667	0.711538462
5	0.559701493	1	0.717703349
6	0.559701493	1	0.717703349
7	0.559701493	1	0.717703349
8	0.559701493	1	0.717703349
9	0.551181102	0.933333333	0.693069307
10	0.56	0.933333333	0.7
11	0.5	0.013333333	0.025974026
13	0.559701493	1	0.717703349
14	1	0.013333333	0.026315789
15	0.2	0.013333333	0.025
16	0.6	0.04	0.075
17	0.559701493	1	0.717703349
18	0.556390977	0.986666667	0.711538462
19	0.559701493	1	0.717703349
20	0.563909774	1	0.721153846
21	0.559701493	1	0.717703349
22	0.556390977	0.986666667	0.711538462
24	0.559701493	1	0.717703349
25	0.559701493	1	0.717703349
26	0.559701493	1	0.717703349
27	0.553030303	0.973333333	0.70531401
28	0.556390977	0.986666667	0.711538462
29	0.559701493	1	0.717703349
30	0.559701493	1	0.717703349
31	1	0.013333333	0.026315789
32	0.556390977	0.986666667	0.711538462
33	0.556390977	0.986666667	0.711538462
34	0.556390977	0.986666667	0.711538462
35	0.556390977	0.986666667	0.711538462
36	0.563909774	1	0.721153846
37	0.556390977	0.986666667	0.711538462
38	0.559701493	1	0.717703349
39	0.559701493	1	0.717703349
40	0.559701493	1	0.717703349
41	0.559701493	1	0.717703349
42	0.559701493	1	0.717703349
43	0.559701493	1	0.717703349
44	0.559701493	1	0.717703349

45	0.559701493	1	0.717703349
46	0.559701493	1	0.717703349
47	0.559701493	1	0.717703349
48	0.563909774	1	0.721153846
49	0.563909774	1	0.721153846
50	0.563909774	1	0.721153846
51	0.559701493	1	0.717703349
52	0.559701493	1	0.717703349
53	0.556390977	0.986666667	0.711538462
54	0.559701493	1	0.717703349
55	0.556390977	0.986666667	0.711538462
56	0.556390977	0.986666667	0.711538462
59	0.559701493	1	0.717703349
60	0.559701493	1	0.717703349
61	0.8	0.053333333	0.1
62	0.5	0.053333333	0.096385542
63	0.559701493	1	0.717703349
64	0.5	0.04	0.074074074
65	0.559701493	1	0.717703349
66	0.559701493	1	0.717703349
67	0.559701493	1	0.717703349
68	0.559701493	1	0.717703349
69	0.559701493	1	0.717703349
71	0.563909774	1	0.721153846
72	0.559701493	1	0.717703349
73	0.556390977	0.986666667	0.711538462
74	0.559701493	1	0.717703349
76	0.559701493	1	0.717703349
77	0.515151515	0.226666667	0.314814815
78	0.596491228	0.453333333	0.515151515
79	0.559701493	1	0.717703349
80	0.559701493	1	0.717703349
81	0.559701493	1	0.717703349
82	0.559701493	1	0.717703349
83	0.559701493	1	0.717703349
84	0.559701493	1	0.717703349

Table G-56. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.440298507	1	0.611398964
1	0.333333333	0.016949153	0.032258065
2	0.444444444	0.06779661	0.117647059
9	0.285714286	0.033898305	0.060606061
10	0.444444444	0.06779661	0.117647059
11	0.439393939	0.983050847	0.607329843
12	0.427480916	0.949152542	0.589473684
14	0.443609023	1	0.614583333
15	0.426356589	0.93220339	0.585106383
16	0.441860465	0.966101695	0.606382979
20	1	0.016949153	0.033333333
23	0.440298507	1	0.611398964
31	0.443609023	1	0.614583333
36	1	0.016949153	0.033333333
48	1	0.016949153	0.033333333
49	1	0.016949153	0.033333333
50	1	0.016949153	0.033333333
57	0.440298507	1	0.611398964
58	0.440298507	1	0.611398964
61	0.449612403	0.983050847	0.617021277
62	0.436507937	0.93220339	0.594594595
64	0.4375	0.949152542	0.598930481
70	0.440298507	1	0.611398964
71	1	0.016949153	0.033333333
75	0.440298507	1	0.611398964
77	0.425742574	0.728813559	0.5375
78	0.467532468	0.610169492	0.529411765

Table G-57. P, R, F-Score for Females

7. Extracted Test Data: 20s and 40s

Male	Precision	Recall	F-Score
Baseline	0.558139535	1	0.71641791
1	0.559055118	0.986111111	0.713567839
2	0.552	0.958333333	0.700507614
3	0.558139535	1	0.71641791
4	0.558139535	1	0.71641791
5	0.558139535	1	0.71641791
6	0.558139535	1	0.71641791
8	0.558139535	1	0.71641791
9	0.552845528	0.944444444	0.697435897
10	0.5546875	0.986111111	0.71
12	0.333333333	0.013888889	0.026666667
13	0.558139535	1	0.71641791
14	0.5546875	0.986111111	0.71
15	0.333333333	0.027777778	0.051282051
16	0.666666667	0.027777778	0.053333333
17	0.558139535	1	0.71641791
18	0.5546875	0.986111111	0.71
19	0.558139535	1	0.71641791
20	0.558139535	1	0.71641791
21	0.558139535	1	0.71641791
22	0.551181102	0.972222222	0.703517588
24	1	0.013888889	0.02739726
25	0.558139535	1	0.71641791
26	0.5546875	0.986111111	0.71
27	0.5546875	0.986111111	0.71
28	1	0.027777778	0.054054054
29	0.558139535	1	0.71641791
30	0.558139535	1	0.71641791
32	0.5546875	0.986111111	0.71
33	0.5546875	0.986111111	0.71
34	0.5546875	0.986111111	0.71
35	0.558139535	1	0.71641791
36	0.5625	1	0.72
39	0.558139535	1	0.71641791
41	0.558139535	1	0.71641791
42	0.558139535	1	0.71641791
43	0.558139535	1	0.71641791
44	0.558139535	1	0.71641791
45	0.558139535	1	0.71641791
48	0.558139535	1	0.71641791
51	0.558139535	1	0.71641791
52	0.558139535	1	0.71641791
53	0.5546875	0.986111111	0.71

55	1	0.013888889	0.02739726
56	0.558139535	1	0.71641791
57	0.558139535	1	0.71641791
59	0.5546875	0.986111111	0.71
60	0.5546875	0.986111111	0.71
61	0.666666667	0.027777778	0.053333333
62	0.727272727	0.111111111	0.192771084
63	0.558139535	1	0.71641791
64	0.5	0.041666667	0.076923077
65	0.558139535	1	0.71641791
66	0.558139535	1	0.71641791
67	0.558139535	1	0.71641791
68	1	0.013888889	0.02739726
69	0.558139535	1	0.71641791
70	1	0.013888889	0.02739726
71	0.558139535	1	0.71641791
72	0.558139535	1	0.71641791
73	0.5546875	0.986111111	0.71
76	0.558139535	1	0.71641791
77	0.545454545	0.25	0.342857143
78	0.675675676	0.347222222	0.458715596
79	0.558139535	1	0.71641791
80	0.558139535	1	0.71641791
81	0.558139535	1	0.71641791
82	0.558139535	1	0.71641791
83	0.558139535	1	0.71641791
84	0.558139535	1	0.71641791

Table G-58. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.441860465	1	0.612903226
1	0.5	0.01754386	0.033898305
2	0.25	0.01754386	0.032786885
7	0.441860465	1	0.612903226
9	0.333333333	0.035087719	0.063492063
11	0.433070866	0.964912281	0.597826087
12	0.436507937	0.964912281	0.601092896
15	0.430894309	0.929824561	0.588888889
16	0.444444444	0.98245614	0.612021858
23	0.441860465	1	0.612903226
24	0.4453125	1	0.616216216
28	0.448818898	1	0.619565217
31	0.441860465	1	0.612903226
36	1	0.01754386	0.034482759
37	0.441860465	1	0.612903226
38	0.441860465	1	0.612903226
40	0.441860465	1	0.612903226
46	0.441860465	1	0.612903226
47	0.441860465	1	0.612903226
49	0.441860465	1	0.612903226
50	0.4375	0.98245614	0.605405405
54	0.441860465	1	0.612903226
55	0.4453125	1	0.616216216
58	0.441860465	1	0.612903226
61	0.444444444	0.98245614	0.612021858
62	0.457627119	0.947368421	0.617142857
64	0.43902439	0.947368421	0.6
68	0.4453125	1	0.616216216
70	0.4453125	1	0.616216216
74	0.441860465	1	0.612903226
75	0.441860465	1	0.612903226
77	0.4375	0.736842105	0.549019608
78	0.489130435	0.789473684	0.604026846

Table G-59. P, R, F-Score for Females

8. Extracted Test Data: 20s and 50s

Male	Precision	Recall	F-Score
Baseline	0.560747664	1	0.718562874
1	0.561904762	0.983333333	0.715151515
2	0.572916667	0.916666667	0.705128205
3	0.560747664	1	0.718562874
4	0.560747664	1	0.718562874
5	0.560747664	1	0.718562874
6	0.560747664	1	0.718562874
7	0.560747664	1	0.718562874
8	0.560747664	1	0.718562874
9	0.554455446	0.933333333	0.695652174
10	0.556603774	0.983333333	0.710843373
11	0.567307692	0.983333333	0.719512195
12	0.5	0.033333333	0.0625
13	0.560747664	1	0.718562874
14	0.556603774	0.983333333	0.710843373
15	0.2	0.016666667	0.030769231
16	0.666666667	0.033333333	0.063492063
17	0.556603774	0.983333333	0.710843373
18	0.556603774	0.983333333	0.710843373
19	0.560747664	1	0.718562874
20	0.560747664	1	0.718562874
21	0.560747664	1	0.718562874
22	0.556603774	0.983333333	0.710843373
23	0.560747664	1	0.718562874
24	1	0.016666667	0.032786885
25	0.560747664	1	0.718562874
26	0.560747664	1	0.718562874
27	0.556603774	0.983333333	0.710843373
28	1	0.033333333	0.064516129
29	0.560747664	1	0.718562874
30	0.560747664	1	0.718562874
31	0.556603774	0.983333333	0.710843373
32	0.556603774	0.983333333	0.710843373
33	0.560747664	1	0.718562874
34	0.556603774	0.983333333	0.710843373
37	0.560747664	1	0.718562874
39	0.560747664	1	0.718562874
40	0.556603774	0.983333333	0.710843373
41	0.560747664	1	0.718562874
42	0.560747664	1	0.718562874
43	0.560747664	1	0.718562874
44	0.560747664	1	0.718562874
45	0.560747664	1	0.718562874

46	0.560747664	1	0.718562874
47	0.560747664	1	0.718562874
48	0.560747664	1	0.718562874
50	1	0.016666667	0.032786885
51	0.560747664	1	0.718562874
52	0.560747664	1	0.718562874
53	0.556603774	0.983333333	0.710843373
54	0.560747664	1	0.718562874
55	0.556603774	0.983333333	0.710843373
56	0.571428571	1	0.727272727
57	0.560747664	1	0.718562874
59	0.560747664	1	0.718562874
60	0.560747664	1	0.718562874
61	0.5	0.016666667	0.032258065
62	0.666666667	0.1	0.173913043
63	0.560747664	1	0.718562874
64	0.5	0.05	0.090909091
65	0.560747664	1	0.718562874
66	0.560747664	1	0.718562874
67	0.560747664	1	0.718562874
68	0.560747664	1	0.718562874
69	0.560747664	1	0.718562874
70	0.560747664	1	0.718562874
71	0.560747664	1	0.718562874
72	0.560747664	1	0.718562874
73	0.556603774	0.983333333	0.710843373
76	0.560747664	1	0.718562874
77	0.5	0.233333333	0.318181818
78	0.62745098	0.533333333	0.576576577
79	0.560747664	1	0.718562874
80	0.560747664	1	0.718562874
81	0.560747664	1	0.718562874
82	0.560747664	1	0.718562874
83	0.560747664	1	0.718562874
84	0.560747664	1	0.718562874

Table G-60. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.439252336	1	0.61038961
1	0.5	0.021276596	0.040816327
2	0.545454545	0.127659574	0.206896552
9	0.333333333	0.042553191	0.075471698
11	0.666666667	0.042553191	0.08
12	0.436893204	0.957446809	0.6
15	0.421568627	0.914893617	0.577181208
16	0.442307692	0.978723404	0.609271523
24	0.443396226	1	0.614379085
28	0.447619048	1	0.618421053
35	0.439252336	1	0.61038961
36	0.439252336	1	0.61038961
38	0.439252336	1	0.61038961
49	0.439252336	1	0.61038961
50	0.443396226	1	0.614379085
56	1	0.042553191	0.081632653
58	0.439252336	1	0.61038961
61	0.438095238	0.978723404	0.605263158
62	0.448979592	0.936170213	0.606896552
64	0.435643564	0.936170213	0.594594595
74	0.439252336	1	0.61038961
75	0.439252336	1	0.61038961
77	0.417721519	0.70212766	0.523809524
78	0.5	0.595744681	0.54368932

Table G-61. P, R, F-Score for Females

9. Extracted Test Data: 30s and 40s

Male	Precision	Recall	F-Score
Baseline	0.623188406	1	0.767857143
1	0.623188406	1	0.767857143
2	0.611940299	0.953488372	0.745454545
3	0.626865672	0.976744186	0.763636364
4	0.626865672	0.976744186	0.763636364
5	0.623188406	1	0.767857143
6	0.623188406	1	0.767857143
7	0.623188406	1	0.767857143
8	0.623188406	1	0.767857143
9	0.615384615	0.930232558	0.740740741
10	0.626865672	0.976744186	0.763636364
11	0.623188406	1	0.767857143
12	0.5	0.023255814	0.044444444
14	0.617647059	0.976744186	0.756756757
15	1	0.046511628	0.088888889
16	1	0.023255814	0.045454545
17	0.623188406	1	0.767857143
18	0.623188406	1	0.767857143
19	0.632352941	1	0.774774775
20	0.632352941	1	0.774774775
21	0.617647059	0.976744186	0.756756757
22	0.617647059	0.976744186	0.756756757
24	0.623188406	1	0.767857143
25	0.623188406	1	0.767857143
26	0.623188406	1	0.767857143
27	0.623188406	1	0.767857143
28	0.623188406	1	0.767857143
31	0.623188406	1	0.767857143
32	0.611940299	0.953488372	0.745454545
33	0.623188406	1	0.767857143
34	0.623188406	1	0.767857143
35	0.623188406	1	0.767857143
36	0.623188406	1	0.767857143
37	0.623188406	1	0.767857143
38	0.623188406	1	0.767857143
39	0.617647059	0.976744186	0.756756757
40	0.617647059	0.976744186	0.756756757
41	0.623188406	1	0.767857143
42	0.623188406	1	0.767857143
43	0.623188406	1	0.767857143
44	0.623188406	1	0.767857143
45	0.632352941	1	0.774774775
46	0.632352941	1	0.774774775

47	0.623188406	1	0.767857143
48	0.623188406	1	0.767857143
50	0.333333333	0.023255814	0.043478261
51	0.623188406	1	0.767857143
52	0.623188406	1	0.767857143
53	0.623188406	1	0.767857143
54	0.623188406	1	0.767857143
55	0.626865672	0.976744186	0.763636364
56	0.617647059	0.976744186	0.756756757
57	0.623188406	1	0.767857143
59	0.617647059	0.976744186	0.756756757
60	0.617647059	0.976744186	0.756756757
61	0.617647059	0.976744186	0.756756757
62	0.625	0.11627907	0.196078431
63	0.623188406	1	0.767857143
65	0.623188406	1	0.767857143
66	0.623188406	1	0.767857143
67	0.623188406	1	0.767857143
68	0.617647059	0.976744186	0.756756757
71	0.632352941	1	0.774774775
72	0.623188406	1	0.767857143
73	0.623188406	1	0.767857143
74	0.623188406	1	0.767857143
75	0.623188406	1	0.767857143
76	0.623188406	1	0.767857143
77	0.642857143	0.209302326	0.315789474
78	0.631578947	0.558139535	0.592592593
79	0.623188406	1	0.767857143
80	0.623188406	1	0.767857143
81	0.623188406	1	0.767857143
82	0.623188406	1	0.767857143
83	0.623188406	1	0.767857143
84	0.623188406	1	0.767857143

Table G-62. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.376811594	1	0.547368421
3	0.5	0.038461538	0.071428571
4	0.5	0.038461538	0.071428571
9	0.25	0.038461538	0.066666667
10	0.5	0.038461538	0.071428571
12	0.373134328	0.961538462	0.537634409
13	0.376811594	1	0.547368421
15	0.388059701	1	0.559139785
16	0.382352941	1	0.553191489
19	1	0.038461538	0.074074074
20	1	0.038461538	0.074074074
23	0.376811594	1	0.547368421
29	0.376811594	1	0.547368421
30	0.376811594	1	0.547368421
45	1	0.038461538	0.074074074
46	1	0.038461538	0.074074074
49	0.367647059	0.961538462	0.531914894
50	0.363636364	0.923076923	0.52173913
55	0.5	0.038461538	0.071428571
58	0.376811594	1	0.547368421
62	0.37704918	0.884615385	0.528735632
64	0.376811594	1	0.547368421
69	0.376811594	1	0.547368421
70	0.376811594	1	0.547368421
71	1	0.038461538	0.074074074
77	0.381818182	0.807692308	0.518518519
78	0.387096774	0.461538462	0.421052632

Table G-63. P, R, F-Score for Females

10. Extracted Test Data: 30s and 50s

Male	Precision	Recall	F-Score
Baseline	0.659574468	1	0.794871795
1	1	0.032258065	0.0625
2	0.673913043	1	0.805194805
3	0.652173913	0.967741935	0.779220779
4	0.659574468	1	0.794871795
5	0.659574468	1	0.794871795
6	0.659574468	1	0.794871795
7	0.659574468	1	0.794871795
8	0.659574468	1	0.794871795
9	0.666666667	0.967741935	0.789473684
10	0.659090909	0.935483871	0.773333333
11	0.5	0.032258065	0.060606061
12	0.659090909	0.935483871	0.773333333
13	0.652173913	0.967741935	0.779220779
14	0.652173913	0.967741935	0.779220779
15	1	0.064516129	0.121212121
16	0.659574468	1	0.794871795
17	0.659574468	1	0.794871795
18	0.659574468	1	0.794871795
19	0.673913043	1	0.805194805
20	0.673913043	1	0.805194805
21	0.659574468	1	0.794871795
22	0.659574468	1	0.794871795
23	0.659574468	1	0.794871795
24	0.659574468	1	0.794871795
25	0.659574468	1	0.794871795
26	0.659574468	1	0.794871795
27	0.659574468	1	0.794871795
28	0.659574468	1	0.794871795
31	0.659574468	1	0.794871795
32	0.659574468	1	0.794871795
33	0.659574468	1	0.794871795
34	0.659574468	1	0.794871795
35	0.652173913	0.967741935	0.779220779
36	0.659574468	1	0.794871795
37	0.652173913	0.967741935	0.779220779
38	0.652173913	0.967741935	0.779220779
39	0.659574468	1	0.794871795
40	0.659574468	1	0.794871795
41	0.659574468	1	0.794871795
42	0.659574468	1	0.794871795
43	0.659574468	1	0.794871795
44	0.659574468	1	0.794871795

45	0.659574468	1	0.794871795
46	0.659574468	1	0.794871795
47	0.681818182	0.967741935	0.8
48	0.673913043	1	0.805194805
49	0.666666667	0.967741935	0.789473684
50	0.673913043	1	0.805194805
51	0.659574468	1	0.794871795
52	0.659574468	1	0.794871795
53	0.659574468	1	0.794871795
54	0.652173913	0.967741935	0.779220779
55	0.659574468	1	0.794871795
56	0.652173913	0.967741935	0.779220779
57	0.659574468	1	0.794871795
58	0.659574468	1	0.794871795
59	0.659574468	1	0.794871795
60	0.659574468	1	0.794871795
61	0.644444444	0.935483871	0.763157895
62	0.666666667	0.903225806	0.767123288
63	0.659574468	1	0.794871795
65	0.659574468	1	0.794871795
66	0.659574468	1	0.794871795
67	0.659574468	1	0.794871795
68	0.659574468	1	0.794871795
69	0.659574468	1	0.794871795
70	0.659574468	1	0.794871795
71	0.673913043	1	0.805194805
72	0.659574468	1	0.794871795
75	0.659574468	1	0.794871795
76	0.659574468	1	0.794871795
77	0.615384615	0.258064516	0.363636364
78	0.666666667	0.193548387	0.3
79	0.659574468	1	0.794871795
80	0.659574468	1	0.794871795
81	0.659574468	1	0.794871795
82	0.659574468	1	0.794871795
83	0.673913043	1	0.805194805
84	0.659574468	1	0.794871795

Table G-64. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.340425532	1	0.507936508
1	0.347826087	1	0.516129032
2	1	0.0625	0.117647059
9	0.5	0.0625	0.111111111
10	0.333333333	0.0625	0.105263158
11	0.333333333	0.9375	0.491803279
12	0.333333333	0.0625	0.105263158
15	0.355555556	1	0.524590164
19	1	0.0625	0.117647059
20	1	0.0625	0.117647059
29	0.340425532	1	0.507936508
30	0.340425532	1	0.507936508
47	0.666666667	0.125	0.210526316
48	1	0.0625	0.117647059
49	0.5	0.0625	0.111111111
50	1	0.0625	0.117647059
62	0.4	0.125	0.19047619
64	0.340425532	1	0.507936508
71	1	0.0625	0.117647059
73	0.340425532	1	0.507936508
74	0.340425532	1	0.507936508
77	0.323529412	0.6875	0.44
78	0.342105263	0.8125	0.481481481
83	1	0.0625	0.117647059

Table G-65. P, R, F-Score for Females

11. Extracted Test Data: 40s and 50s

Male	Precision	Recall	F-Score
Baseline	0.666666667	1	0.8
1	0.666666667	1	0.8
2	0.666666667	1	0.8
3	0.682926829	1	0.811594203
4	0.682926829	1	0.811594203
5	0.666666667	1	0.8
6	0.682926829	1	0.811594203
7	0.666666667	1	0.8
8	0.666666667	1	0.8
9	0.65	0.928571429	0.764705882
10	0.658536585	0.964285714	0.782608696
12	0.641025641	0.892857143	0.746268657
13	0.666666667	1	0.8
14	0.666666667	1	0.8
15	0.658536585	0.964285714	0.782608696
16	0.666666667	1	0.8
17	0.666666667	1	0.8
18	0.666666667	1	0.8
19	0.666666667	1	0.8
20	0.666666667	1	0.8
21	0.666666667	1	0.8
22	0.666666667	1	0.8
23	0.666666667	1	0.8
24	0.666666667	1	0.8
25	0.666666667	1	0.8
26	0.666666667	1	0.8
27	0.658536585	0.964285714	0.782608696
28	0.658536585	0.964285714	0.782608696
29	0.666666667	1	0.8
30	0.666666667	1	0.8
31	0.666666667	1	0.8
32	0.666666667	1	0.8
33	0.666666667	1	0.8
34	0.666666667	1	0.8
35	0.666666667	1	0.8
36	0.666666667	1	0.8
37	0.666666667	1	0.8
38	0.666666667	1	0.8
39	0.658536585	0.964285714	0.782608696
40	0.658536585	0.964285714	0.782608696
41	0.666666667	1	0.8
42	0.666666667	1	0.8
43	0.666666667	1	0.8

44	0.666666667	1	0.8
45	0.682926829	1	0.811594203
46	0.682926829	1	0.811594203
47	0.666666667	1	0.8
48	0.666666667	1	0.8
49	0.666666667	1	0.8
50	0.666666667	1	0.8
51	0.658536585	0.964285714	0.782608696
52	0.658536585	0.964285714	0.782608696
53	0.666666667	1	0.8
54	0.666666667	1	0.8
55	0.666666667	1	0.8
56	0.666666667	1	0.8
57	0.666666667	1	0.8
58	0.666666667	1	0.8
59	0.666666667	1	0.8
60	0.658536585	0.964285714	0.782608696
61	0.666666667	1	0.8
62	1	0.071428571	0.133333333
63	0.666666667	1	0.8
65	0.666666667	1	0.8
66	0.666666667	1	0.8
67	0.666666667	1	0.8
68	0.666666667	1	0.8
69	0.666666667	1	0.8
70	0.675	0.964285714	0.794117647
71	0.666666667	1	0.8
72	0.666666667	1	0.8
73	0.666666667	1	0.8
74	0.666666667	1	0.8
75	0.666666667	1	0.8
76	0.666666667	1	0.8
77	0.666666667	1	0.8
78	0.733333333	0.392857143	0.511627907
79	0.666666667	1	0.8
80	0.666666667	1	0.8
81	0.666666667	1	0.8
82	0.666666667	1	0.8
83	0.666666667	1	0.8
84	0.658536585	0.964285714	0.782608696

Table G-66.

P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.333333333	1	0.5
3	1	0.071428571	0.133333333
4	1	0.071428571	0.133333333
6	1	0.071428571	0.133333333
11	0.333333333	1	0.5
45	1	0.071428571	0.133333333
46	1	0.071428571	0.133333333
62	0.35	1	0.518518519
64	0.333333333	1	0.5
70	0.5	0.071428571	0.125
78	0.37037037	0.714285714	0.487804878

Table G-67. P, R, F-Score for Females

12. Extracted Test Data: Under 26 and 26 or Over

Male	Precision	Recall	F-Score
Baseline	0.512295082	1	0.677506775
2	0.522123894	0.944	0.672364672
3	0.510288066	0.992	0.673913043
4	0.510288066	0.992	0.673913043
5	0.512295082	1	0.677506775
6	0.512295082	1	0.677506775
7	0.512295082	1	0.677506775
8	0.512295082	1	0.677506775
9	0.514403292	1	0.679347826
10	0.514893617	0.968	0.672222222
11	0.25	0.008	0.015503876
12	0.444444444	0.032	0.059701493
14	0.510373444	0.984	0.672131148
15	0.307692308	0.032	0.057971014
16	0.583333333	0.056	0.102189781
17	1	0.008	0.015873016
18	0.518672199	1	0.683060109
19	0.512295082	1	0.677506775
21	0.512295082	1	0.677506775
22	0.510288066	0.992	0.673913043
23	0.512295082	1	0.677506775
24	0.512295082	1	0.677506775
25	0.512295082	1	0.677506775
27	0.510288066	0.992	0.673913043
28	1	0.008	0.015873016
30	0.512295082	1	0.677506775
31	0.510288066	0.992	0.673913043
32	0.510288066	0.992	0.673913043
33	1	0.008	0.015873016
34	0.5	0.008	0.015748031
35	0.514522822	0.992	0.677595628
37	0.5	0.008	0.015748031
38	0.5	0.008	0.015748031
39	1	0.016	0.031496063
40	1	0.008	0.015873016
42	0.512295082	1	0.677506775
43	0.512295082	1	0.677506775
45	0.512295082	1	0.677506775
50	0.514403292	1	0.679347826
51	0.512295082	1	0.677506775
52	0.512295082	1	0.677506775
53	0.512295082	1	0.677506775
55	0.5	0.008	0.015748031

56	0.521008403	0.992	0.683195592
59	0.512295082	1	0.677506775
62	0.4	0.016	0.030769231
63	0.512295082	1	0.677506775
64	0.416666667	0.04	0.072992701
65	0.512295082	1	0.677506775
66	0.512295082	1	0.677506775
67	0.512295082	1	0.677506775
68	0.512295082	1	0.677506775
71	0.512295082	1	0.677506775
72	0.512295082	1	0.677506775
73	0.512295082	1	0.677506775
76	0.512295082	1	0.677506775
77	0.406779661	0.192	0.260869565
78	0.53968254	0.272	0.361702128
79	0.512295082	1	0.677506775
80	0.512295082	1	0.677506775
81	0.512295082	1	0.677506775
82	0.512295082	1	0.677506775

Table G-68. P, R, F-Score for Males

Female	Precision	Recall	F-Score
Baseline	0.487704918	1	0.655647383
1	0.487704918	1	0.655647383
2	0.611111111	0.092436975	0.160583942
9	1	0.008403361	0.016666667
10	0.555555556	0.042016807	0.078125
11	0.483333333	0.974789916	0.646239554
12	0.485106383	0.957983193	0.644067797
13	0.487704918	1	0.655647383
14	0.333333333	0.008403361	0.016393443
15	0.476190476	0.924369748	0.628571429
16	0.49137931	0.957983193	0.64957265
17	0.489711934	1	0.657458564
18	1	0.025210084	0.049180328
20	0.487704918	1	0.655647383
26	0.487704918	1	0.655647383
28	0.489711934	1	0.657458564
29	0.487704918	1	0.655647383
33	0.489711934	1	0.657458564
34	0.487603306	0.991596639	0.653739612
35	0.666666667	0.016806723	0.032786885
36	0.487704918	1	0.655647383
37	0.487603306	0.991596639	0.653739612
38	0.487603306	0.991596639	0.653739612
39	0.491735537	1	0.659279778
40	0.489711934	1	0.657458564
41	0.487704918	1	0.655647383
44	0.487704918	1	0.655647383
46	0.487704918	1	0.655647383
47	0.487704918	1	0.655647383
48	0.485596708	0.991596639	0.651933702
49	0.487704918	1	0.655647383
50	1	0.008403361	0.016666667
54	0.487704918	1	0.655647383
55	0.487603306	0.991596639	0.653739612
56	0.833333333	0.042016807	0.08
57	0.487704918	1	0.655647383
58	0.487704918	1	0.655647383
60	0.487704918	1	0.655647383
61	0.487704918	1	0.655647383
62	0.485355649	0.974789916	0.648044693
64	0.482758621	0.941176471	0.638176638
69	0.487704918	1	0.655647383
70	0.487704918	1	0.655647383

74	0.487704918	1	0.655647383
75	0.487704918	1	0.655647383
77	0.454054054	0.705882353	0.552631579
78	0.497237569	0.756302521	0.6
83	0.487704918	1	0.655647383
84	0.487704918	1	0.655647383

Table G-69. P, R, F-Score for Females

E. AGE: MULTI-CLASS (5-WAY) CLASSIFICATION WITHOUT PRIOR

1. All Test Data

13-19	Precision	Recall	F-Score
Baseline	0.278688525	1	0.435897436
2	0.333333333	0.029411765	0.054054054
9	0.280851064	0.970588235	0.435643564
10	0.284482759	0.970588235	0.44
11	0.280334728	0.985294118	0.436482085
12	0.278969957	0.955882353	0.431893688
16	0.666666667	0.058823529	0.108108108
48	0.275720165	0.985294118	0.430868167
54	0.278688525	1	0.435897436
62	0.272727273	0.044117647	0.075949367
63	0.275720165	0.985294118	0.430868167
74	0.278688525	1	0.435897436
77	0.355932203	0.308823529	0.330708661

Table G-70. P, R, F-Score for Teens

20-29	Precision	Recall	F-Score
Baseline	0.397540984	1	0.568914956
1	0.395833333	0.979381443	0.56379822
3	0.399176955	1	0.570588235
4	0.399176955	1	0.570588235
5	0.397540984	1	0.568914956
6	0.397540984	1	0.568914956
7	0.397540984	1	0.568914956
8	0.397540984	1	0.568914956
9	0.75	0.06185567	0.114285714
10	0.666666667	0.06185567	0.113207547
11	0.5	0.020618557	0.03960396
13	0.397540984	1	0.568914956
14	0.40167364	0.989690722	0.571428571
15	0.5	0.030927835	0.058252427
17	0.395061728	0.989690722	0.564705882
18	0.4	0.989690722	0.569732938
19	0.397540984	1	0.568914956
20	0.399176955	1	0.570588235
21	0.397540984	1	0.568914956
22	0.395061728	0.989690722	0.564705882
23	0.399176955	1	0.570588235
25	0.399176955	1	0.570588235
26	0.399176955	1	0.570588235
27	0.392561983	0.979381443	0.560471976
28	0.392561983	0.979381443	0.560471976
29	0.399176955	1	0.570588235
30	0.397540984	1	0.568914956
31	0.395061728	0.989690722	0.564705882
32	0.395061728	0.989690722	0.564705882
33	0.396694215	0.989690722	0.566371681
34	0.398340249	0.989690722	0.568047337
35	0.402489627	1	0.573964497
36	0.397540984	1	0.568914956
37	0.402489627	1	0.573964497
38	0.402489627	1	0.573964497
39	0.396694215	0.989690722	0.566371681
40	0.399176955	1	0.570588235
41	0.397540984	1	0.568914956
42	0.399176955	1	0.570588235
43	0.397540984	1	0.568914956
44	0.397540984	1	0.568914956
45	0.397540984	1	0.568914956
46	0.397540984	1	0.568914956

47	0.399176955	1	0.570588235
49	0.397540984	1	0.568914956
50	0.399176955	1	0.570588235
51	0.397540984	1	0.568914956
52	0.397540984	1	0.568914956
53	0.395061728	0.989690722	0.564705882
55	0.396694215	0.989690722	0.566371681
57	0.399176955	1	0.570588235
58	0.397540984	1	0.568914956
59	0.397540984	1	0.568914956
60	0.399176955	1	0.570588235
61	0.397540984	1	0.568914956
64	0.5	0.06185567	0.110091743
65	0.397540984	1	0.568914956
66	0.397540984	1	0.568914956
67	0.397540984	1	0.568914956
68	0.397540984	1	0.568914956
69	0.399176955	1	0.570588235
70	0.399176955	1	0.570588235
71	0.399176955	1	0.570588235
72	0.397540984	1	0.568914956
73	0.395061728	0.989690722	0.564705882
75	0.397540984	1	0.568914956
76	0.397540984	1	0.568914956
78	0.454545455	0.257731959	0.328947368
79	0.397540984	1	0.568914956
80	0.397540984	1	0.568914956
81	0.397540984	1	0.568914956
82	0.397540984	1	0.568914956
83	0.397540984	1	0.568914956
84	0.397540984	1	0.568914956

Table G-71. P, R, F-Score for 20s

30-39	Precision	Recall	F-Score
Baseline	0.151639344	1	0.263345196
2	0.163716814	1	0.281368821
12	0.25	0.027027027	0.048780488
24	0.152892562	1	0.265232975
56	0.151260504	0.972972973	0.261818182
78	0.206349206	0.702702703	0.319018405

Table G-72. P, R, F-Score for 30s

40-49	Precision	Recall	F-Score
Baseline	0.131147541	1	0.231884058
15	0.134782609	0.96875	0.236641221
16	0.139130435	1	0.244274809
62	0.135746606	0.9375	0.23715415
77	0.152542373	0.84375	0.258373206

Table G-73. P, R, F-Score for 40s

50-59	Precision	Recall	F-Score
Baseline	0.040983607	1	0.078740157
12	0.285714286	0.2	0.235294118
64	0.043103448	1	0.082644628
78	0.047619048	0.3	0.082191781

Table G-74. P, R, F-Score for 50s

F. AGE: BINARY CLASSIFICATION WITHOUT PRIOR

1. Extracted Test Data: Teens and 20s

13-19	Precision	Recall	F-Score
Baseline	0.412121212	1	0.583690987
2	0.333333333	0.073529412	0.120481928
9	0.424050633	0.985294118	0.592920354
10	0.423076923	0.970588235	0.589285714
11	0.40625	0.955882353	0.570175439
12	0.41875	0.985294118	0.587719298
14	0.666666667	0.029411765	0.056338028
15	0.571428571	0.058823529	0.106666667
16	0.714285714	0.073529412	0.133333333
17	0.666666667	0.029411765	0.056338028
18	0.75	0.044117647	0.083333333
23	1	0.014705882	0.028985507
24	0.411042945	0.985294118	0.58008658
29	1	0.014705882	0.028985507
32	0.414634146	1	0.586206897
33	0.408536585	0.985294118	0.577586207
34	0.411042945	0.985294118	0.58008658
35	1	0.029411765	0.057142857
37	1	0.014705882	0.028985507
38	1	0.014705882	0.028985507
40	0.412121212	1	0.583690987
42	1	0.014705882	0.028985507
47	1	0.014705882	0.028985507
48	0.408536585	0.985294118	0.577586207
54	0.412121212	1	0.583690987
55	0.5	0.014705882	0.028571429
56	1	0.014705882	0.028985507
57	1	0.014705882	0.028985507
58	0.412121212	1	0.583690987
60	0.412121212	1	0.583690987
62	0.333333333	0.044117647	0.077922078
63	0.412121212	1	0.583690987
64	0.405228758	0.911764706	0.561085973
69	1	0.014705882	0.028985507
70	1	0.014705882	0.028985507
71	1	0.014705882	0.028985507
74	0.412121212	1	0.583690987
77	0.5	0.382352941	0.433333333
78	0.421686747	0.514705882	0.463576159

Table G-75. P, R, F-Score for Teens

20-29	Precision	Recall	F-Score
Baseline	0.587878788	1	0.740458015
1	0.587878788	1	0.740458015
2	0.58	0.896907216	0.704453441
3	0.587878788	1	0.740458015
4	0.587878788	1	0.740458015
5	0.587878788	1	0.740458015
6	0.587878788	1	0.740458015
7	0.587878788	1	0.740458015
8	0.587878788	1	0.740458015
9	0.857142857	0.06185567	0.115384615
10	0.777777778	0.072164948	0.132075472
11	0.4	0.020618557	0.039215686
12	0.8	0.041237113	0.078431373
13	0.587878788	1	0.740458015
14	0.592592593	0.989690722	0.741312741
15	0.594936709	0.969072165	0.737254902
16	0.601265823	0.979381443	0.745098039
17	0.592592593	0.989690722	0.741312741
18	0.596273292	0.989690722	0.744186047
19	0.587878788	1	0.740458015
20	0.587878788	1	0.740458015
21	0.587878788	1	0.740458015
22	0.585365854	0.989690722	0.735632184
23	0.591463415	1	0.743295019
24	0.5	0.010309278	0.02020202
25	0.587878788	1	0.740458015
26	0.587878788	1	0.740458015
27	0.585365854	0.989690722	0.735632184
28	0.585365854	0.989690722	0.735632184
29	0.591463415	1	0.743295019
30	0.587878788	1	0.740458015
31	0.587878788	1	0.740458015
32	1	0.010309278	0.020408163
34	0.5	0.010309278	0.02020202
35	0.595092025	1	0.746153846
36	0.587878788	1	0.740458015
37	0.591463415	1	0.743295019
38	0.591463415	1	0.743295019
39	0.587878788	1	0.740458015
41	0.587878788	1	0.740458015
42	0.591463415	1	0.743295019
43	0.587878788	1	0.740458015
44	0.587878788	1	0.740458015

45	0.587878788	1	0.740458015
46	0.587878788	1	0.740458015
47	0.591463415	1	0.743295019
49	0.587878788	1	0.740458015
50	0.587878788	1	0.740458015
51	0.587878788	1	0.740458015
52	0.587878788	1	0.740458015
53	0.585365854	0.989690722	0.735632184
55	0.588957055	0.989690722	0.738461538
56	0.591463415	1	0.743295019
57	0.591463415	1	0.743295019
59	0.587878788	1	0.740458015
61	0.587878788	1	0.740458015
62	0.583333333	0.93814433	0.719367589
64	0.5	0.06185567	0.110091743
65	0.587878788	1	0.740458015
66	0.587878788	1	0.740458015
67	0.587878788	1	0.740458015
68	0.587878788	1	0.740458015
69	0.591463415	1	0.743295019
70	0.591463415	1	0.743295019
71	0.591463415	1	0.743295019
72	0.587878788	1	0.740458015
73	0.585365854	0.989690722	0.735632184
75	0.587878788	1	0.740458015
76	0.587878788	1	0.740458015
77	0.628318584	0.731958763	0.676190476
78	0.597560976	0.505154639	0.547486034
79	0.587878788	1	0.740458015
80	0.587878788	1	0.740458015
81	0.587878788	1	0.740458015
82	0.587878788	1	0.740458015
83	0.587878788	1	0.740458015
84	0.587878788	1	0.740458015

Table G-76.

P, R, F-Score for 20s

2. Extracted Test Data: Teens and 30s

13-19	Precision	Recall	F-Score
Baseline	0.647619048	1	0.786127168
1	0.647619048	1	0.786127168
2	1	0.073529412	0.136986301
3	0.653846154	1	0.790697674
4	0.653846154	1	0.790697674
5	0.647619048	1	0.786127168
6	0.647619048	1	0.786127168
8	0.647619048	1	0.786127168
9	0.66	0.970588235	0.785714286
10	0.653465347	0.970588235	0.781065089
11	0.643564356	0.955882353	0.769230769
12	0.643564356	0.955882353	0.769230769
13	0.5	0.014705882	0.028571429
14	0.5	0.014705882	0.028571429
15	1	0.132352941	0.233766234
16	0.9	0.132352941	0.230769231
17	0.647619048	1	0.786127168
18	0.647619048	1	0.786127168
19	0.653846154	1	0.790697674
20	0.653846154	1	0.790697674
21	0.647619048	1	0.786127168
22	0.647619048	1	0.786127168
23	0.644230769	0.985294118	0.779069767
24	1	0.014705882	0.028985507
25	0.644230769	0.985294118	0.779069767
26	0.647619048	1	0.786127168
27	0.647619048	1	0.786127168
28	0.647619048	1	0.786127168
29	0.644230769	0.985294118	0.779069767
30	0.644230769	0.985294118	0.779069767
31	0.647619048	1	0.786127168
32	0.647619048	1	0.786127168
33	0.644230769	0.985294118	0.779069767
34	0.640776699	0.970588235	0.771929825
35	0.650485437	0.985294118	0.783625731
36	0.647619048	1	0.786127168
37	0.650485437	0.985294118	0.783625731
38	0.650485437	0.985294118	0.783625731
39	0.647619048	1	0.786127168
40	0.647619048	1	0.786127168
41	0.647619048	1	0.786127168
42	0.644230769	0.985294118	0.779069767
43	0.647619048	1	0.786127168

44	0.647619048	1	0.786127168
45	0.647619048	1	0.786127168
46	0.644230769	0.985294118	0.779069767
47	0.644230769	0.985294118	0.779069767
48	0.644230769	0.985294118	0.779069767
49	0.647619048	1	0.786127168
50	0.653846154	1	0.790697674
51	0.647619048	1	0.786127168
52	0.647619048	1	0.786127168
54	0.647619048	1	0.786127168
55	0.6	0.044117647	0.082191781
56	0.75	0.044117647	0.083333333
57	0.644230769	0.985294118	0.779069767
58	0.647619048	1	0.786127168
59	0.647619048	1	0.786127168
60	0.647619048	1	0.786127168
61	0.647619048	1	0.786127168
62	0.5	0.088235294	0.15
63	0.644230769	0.985294118	0.779069767
64	0.637254902	0.955882353	0.764705882
65	0.647619048	1	0.786127168
66	0.647619048	1	0.786127168
67	0.647619048	1	0.786127168
68	0.647619048	1	0.786127168
69	0.644230769	0.985294118	0.779069767
70	0.644230769	0.985294118	0.779069767
71	0.653846154	1	0.790697674
72	0.647619048	1	0.786127168
73	0.647619048	1	0.786127168
74	0.647619048	1	0.786127168
75	0.647619048	1	0.786127168
76	0.647619048	1	0.786127168
77	0.722222222	0.382352941	0.5
78	0.75	0.485294118	0.589285714

Table G-77.

P, R, F-Score for Teens

30-39	Precision	Recall	F-Score
Baseline	0.352380952	1	0.521126761
2	0.37	1	0.540145985
3	1	0.027027027	0.052631579
4	1	0.027027027	0.052631579
7	0.352380952	1	0.521126761
9	0.6	0.081081081	0.142857143
10	0.5	0.054054054	0.097560976
11	0.25	0.027027027	0.048780488
12	0.25	0.027027027	0.048780488
13	0.349514563	0.972972973	0.514285714
14	0.349514563	0.972972973	0.514285714
15	0.385416667	1	0.556390977
16	0.378947368	0.972972973	0.545454545
19	1	0.027027027	0.052631579
20	1	0.027027027	0.052631579
24	0.355769231	1	0.524822695
35	0.5	0.027027027	0.051282051
37	0.5	0.027027027	0.051282051
38	0.5	0.027027027	0.051282051
50	1	0.027027027	0.052631579
53	0.339805825	0.945945946	0.5
55	0.35	0.945945946	0.510948905
56	0.356435644	0.972972973	0.52173913
62	0.333333333	0.837837838	0.476923077
71	1	0.027027027	0.052631579
77	0.391304348	0.72972973	0.509433962
78	0.426229508	0.702702703	0.530612245

Table G-78. P, R, F-Score for 30s

3. Extracted Test Data: Teens and 40s

13-19	Precision	Recall	F-Score
Baseline	0.68	1	0.80952381
1	0.670103093	0.955882353	0.787878788
2	0.666666667	0.058823529	0.108108108
3	0.68	1	0.80952381
4	0.686868687	1	0.814371257
5	0.68	1	0.80952381
6	0.68	1	0.80952381
7	0.68	1	0.80952381
8	0.68	1	0.80952381
9	1	0.014705882	0.028985507
10	0.680412371	0.970588235	0.8
11	0.676767677	0.985294118	0.80239521
12	0.677083333	0.955882353	0.792682927
13	0.68	1	0.80952381
14	0.676767677	0.985294118	0.80239521
15	0.9	0.132352941	0.230769231
16	1	0.132352941	0.233766234
17	0.68	1	0.80952381
18	0.68	1	0.80952381
19	0.68	1	0.80952381
20	0.68	1	0.80952381
21	0.68	1	0.80952381
22	0.68	1	0.80952381
23	0.676767677	0.985294118	0.80239521
24	0.676767677	0.985294118	0.80239521
25	0.676767677	0.985294118	0.80239521
26	0.683673469	0.985294118	0.807228916
27	0.68	1	0.80952381
28	0.68	1	0.80952381
29	0.676767677	0.985294118	0.80239521
30	0.676767677	0.985294118	0.80239521
31	0.68	1	0.80952381
32	0.68	1	0.80952381
33	0.676767677	0.985294118	0.80239521
34	0.673469388	0.970588235	0.795180723
35	0.676767677	0.985294118	0.80239521
36	0.68	1	0.80952381
37	0.673469388	0.970588235	0.795180723
38	0.673469388	0.970588235	0.795180723
40	0.686868687	1	0.814371257
41	0.68	1	0.80952381
42	0.676767677	0.985294118	0.80239521
43	0.68	1	0.80952381

44	0.68	1	0.80952381
45	0.68	1	0.80952381
46	0.686868687	1	0.814371257
47	0.68	1	0.80952381
48	0.676767677	0.985294118	0.80239521
49	0.68	1	0.80952381
50	0.68	1	0.80952381
51	0.68	1	0.80952381
52	0.68	1	0.80952381
53	0.68	1	0.80952381
54	0.68	1	0.80952381
55	1	0.014705882	0.028985507
56	0.670103093	0.955882353	0.787878788
57	0.676767677	0.985294118	0.80239521
58	0.68	1	0.80952381
59	0.686868687	1	0.814371257
60	0.686868687	1	0.814371257
61	0.68	1	0.80952381
62	1	0.044117647	0.084507042
63	0.676767677	0.985294118	0.80239521
64	1	0.088235294	0.162162162
65	0.68	1	0.80952381
66	0.68	1	0.80952381
67	0.68	1	0.80952381
68	0.686868687	1	0.814371257
69	0.676767677	0.985294118	0.80239521
70	0.676767677	0.985294118	0.80239521
71	0.676767677	0.985294118	0.80239521
72	0.68	1	0.80952381
73	0.68	1	0.80952381
74	0.68	1	0.80952381
75	0.68	1	0.80952381
76	0.68	1	0.80952381
77	0.72972973	0.397058824	0.514285714
78	0.720588235	0.720588235	0.720588235
79	0.68	1	0.80952381
80	0.68	1	0.80952381
81	0.68	1	0.80952381
82	0.68	1	0.80952381
83	0.68	1	0.80952381
84	0.68	1	0.80952381

Table G-79. P, R, F-Score for Teens

40-49	Precision	Recall	F-Score
Baseline	0.32	1	0.484848485
2	0.319148936	0.9375	0.476190476
4	1	0.03125	0.060606061
9	0.323232323	1	0.488549618
10	0.333333333	0.03125	0.057142857
12	0.25	0.03125	0.055555556
15	0.344444444	0.96875	0.508196721
16	0.351648352	1	0.520325203
26	0.5	0.03125	0.058823529
39	0.32	1	0.484848485
40	1	0.03125	0.060606061
46	1	0.03125	0.060606061
55	0.323232323	1	0.488549618
59	1	0.03125	0.060606061
60	1	0.03125	0.060606061
62	0.329896907	1	0.496124031
64	0.340425532	1	0.507936508
68	1	0.03125	0.060606061
77	0.349206349	0.6875	0.463157895
78	0.40625	0.40625	0.40625

Table G-80.

P, R, F-Score for 40s

4. Extracted Test Data: Teens and 50s

13-19	Precision	Recall	F-Score
Baseline	0.871794872	1	0.931506849
1	0.866666667	0.955882353	0.909090909
2	0.868421053	0.970588235	0.916666667
3	0.871794872	1	0.931506849
4	0.871794872	1	0.931506849
5	0.871794872	1	0.931506849
6	0.871794872	1	0.931506849
7	0.871794872	1	0.931506849
8	0.871794872	1	0.931506849
9	0.866666667	0.955882353	0.909090909
10	0.871794872	1	0.931506849
11	0.878378378	0.955882353	0.915492958
12	0.893333333	0.985294118	0.937062937
13	0.871794872	1	0.931506849
14	0.868421053	0.970588235	0.916666667
15	0.863013699	0.926470588	0.893617021
16	0.863013699	0.926470588	0.893617021
17	0.871794872	1	0.931506849
18	0.866666667	0.955882353	0.909090909
19	0.871794872	1	0.931506849
20	0.871794872	1	0.931506849
21	0.871794872	1	0.931506849
22	0.871794872	1	0.931506849
23	0.87012987	0.985294118	0.924137931
24	0.87012987	0.985294118	0.924137931
25	0.87012987	0.985294118	0.924137931
26	0.87012987	0.985294118	0.924137931
27	0.871794872	1	0.931506849
28	0.871794872	1	0.931506849
29	0.87012987	0.985294118	0.924137931
30	0.87012987	0.985294118	0.924137931
31	0.871794872	1	0.931506849
32	0.871794872	1	0.931506849
33	0.87012987	0.985294118	0.924137931
34	0.868421053	0.970588235	0.916666667
35	0.868421053	0.970588235	0.916666667
36	0.871794872	1	0.931506849
37	0.868421053	0.970588235	0.916666667
38	0.868421053	0.970588235	0.916666667
39	0.871794872	1	0.931506849
40	0.871794872	1	0.931506849
41	0.871794872	1	0.931506849
42	0.87012987	0.985294118	0.924137931

43	0.871794872	1	0.931506849
44	0.871794872	1	0.931506849
45	0.871794872	1	0.931506849
46	0.87012987	0.985294118	0.924137931
47	0.87012987	0.985294118	0.924137931
48	0.87012987	0.985294118	0.924137931
49	0.871794872	1	0.931506849
50	0.871794872	1	0.931506849
51	0.871794872	1	0.931506849
52	0.883116883	1	0.937931034
53	0.871794872	1	0.931506849
54	0.871794872	1	0.931506849
55	0.87012987	0.985294118	0.924137931
56	0.866666667	0.955882353	0.909090909
57	0.868421053	0.970588235	0.916666667
58	0.871794872	1	0.931506849
59	0.871794872	1	0.931506849
60	0.871794872	1	0.931506849
61	0.871794872	1	0.931506849
62	1	0.044117647	0.084507042
63	0.87012987	0.985294118	0.924137931
64	1	0.088235294	0.162162162
65	0.871794872	1	0.931506849
66	0.871794872	1	0.931506849
67	0.871794872	1	0.931506849
68	0.871794872	1	0.931506849
69	0.87012987	0.985294118	0.924137931
70	0.87012987	0.985294118	0.924137931
71	0.87012987	0.985294118	0.924137931
72	0.871794872	1	0.931506849
73	0.871794872	1	0.931506849
74	0.871794872	1	0.931506849
75	0.871794872	1	0.931506849
76	0.871794872	1	0.931506849
77	0.807692308	0.308823529	0.446808511
78	0.875	0.720588235	0.790322581
79	0.871794872	1	0.931506849
80	0.871794872	1	0.931506849
81	0.871794872	1	0.931506849
82	0.871794872	1	0.931506849
83	0.871794872	1	0.931506849
84	0.871794872	1	0.931506849

Table G-81. P, R, F-Score for Teens

50-59	Precision	Recall	F-Score
Baseline	0.128205128	1	0.227272727
11	0.25	0.1	0.142857143
12	0.666666667	0.2	0.307692308
52	1	0.1	0.181818182
62	0.133333333	1	0.235294118
64	0.138888889	1	0.243902439
77	0.096153846	0.5	0.161290323
78	0.136363636	0.3	0.1875

Table G-82.

P, R, F-Score for 50s

5. Extracted Test Data: 20s and 30s

20-29	Precision	Recall	F-Score
Baseline	0.723880597	1	0.83982684
1	0.723880597	1	0.83982684
2	1	0.092783505	0.169811321
3	0.729323308	1	0.843478261
4	0.729323308	1	0.843478261
5	0.723880597	1	0.83982684
6	0.723880597	1	0.83982684
7	0.723880597	1	0.83982684
8	0.723880597	1	0.83982684
9	0.723880597	1	0.83982684
10	0.723880597	1	0.83982684
11	0.723076923	0.969072165	0.828193833
12	0.729323308	1	0.843478261
13	0.723880597	1	0.83982684
14	0.727272727	0.989690722	0.838427948
15	0.721804511	0.989690722	0.834782609
16	0.8	0.041237113	0.078431373
17	0.721804511	0.989690722	0.834782609
18	0.721804511	0.989690722	0.834782609
19	0.723880597	1	0.83982684
20	0.729323308	1	0.843478261
21	0.723880597	1	0.83982684
22	0.721804511	0.989690722	0.834782609
23	0.723880597	1	0.83982684
24	1	0.010309278	0.020408163
25	0.723880597	1	0.83982684
26	0.723880597	1	0.83982684
27	0.71969697	0.979381443	0.829694323
28	0.721804511	0.989690722	0.834782609
29	0.723880597	1	0.83982684
30	0.723880597	1	0.83982684
31	0.721804511	0.989690722	0.834782609
32	0.723880597	1	0.83982684
33	0.721804511	0.989690722	0.834782609
34	0.721804511	0.989690722	0.834782609
35	0.729323308	1	0.843478261
36	0.721804511	0.989690722	0.834782609
37	0.729323308	1	0.843478261
38	0.729323308	1	0.843478261
39	0.721804511	0.989690722	0.834782609
40	0.723880597	1	0.83982684
41	0.723880597	1	0.83982684
42	0.723880597	1	0.83982684

43	0.723880597	1	0.83982684
44	0.723880597	1	0.83982684
45	0.723880597	1	0.83982684
46	0.723880597	1	0.83982684
47	0.723880597	1	0.83982684
48	0.729323308	1	0.843478261
49	0.723880597	1	0.83982684
50	0.729323308	1	0.843478261
51	0.723880597	1	0.83982684
52	0.723880597	1	0.83982684
53	0.721804511	0.989690722	0.834782609
54	0.723880597	1	0.83982684
55	0.721804511	0.989690722	0.834782609
56	1	0.020618557	0.04040404
57	0.723880597	1	0.83982684
58	0.723880597	1	0.83982684
59	0.723880597	1	0.83982684
60	0.723880597	1	0.83982684
61	0.728682171	0.969072165	0.831858407
62	0.444444444	0.041237113	0.075471698
63	0.723880597	1	0.83982684
64	1	0.06185567	0.116504854
65	0.723880597	1	0.83982684
66	0.723880597	1	0.83982684
67	0.723880597	1	0.83982684
68	0.723880597	1	0.83982684
69	0.723880597	1	0.83982684
70	0.723880597	1	0.83982684
71	0.729323308	1	0.843478261
72	0.723880597	1	0.83982684
73	0.721804511	0.989690722	0.834782609
74	0.723880597	1	0.83982684
75	0.723880597	1	0.83982684
76	0.723880597	1	0.83982684
77	0.714285714	0.257731959	0.378787879
78	0.816666667	0.505154639	0.624203822
79	0.723880597	1	0.83982684
80	0.723880597	1	0.83982684
81	0.723880597	1	0.83982684
82	0.723880597	1	0.83982684
83	0.723880597	1	0.83982684
84	0.723880597	1	0.83982684

Table G-83. P, R, F-Score for 20s

30-39	Precision	Recall	F-Score
Baseline	0.276119403	1	0.432748538
2	0.296	1	0.456790123
3	1	0.027027027	0.052631579
4	1	0.027027027	0.052631579
11	0.25	0.027027027	0.048780488
12	1	0.027027027	0.052631579
14	0.5	0.027027027	0.051282051
16	0.279069767	0.972972973	0.43373494
20	1	0.027027027	0.052631579
24	0.278195489	1	0.435294118
35	1	0.027027027	0.052631579
37	1	0.027027027	0.052631579
38	1	0.027027027	0.052631579
48	1	0.027027027	0.052631579
50	1	0.027027027	0.052631579
56	0.28030303	1	0.437869822
61	0.4	0.054054054	0.095238095
62	0.256	0.864864865	0.395061728
64	0.2890625	1	0.448484848
71	1	0.027027027	0.052631579
77	0.272727273	0.72972973	0.397058824
78	0.351351351	0.702702703	0.468468468

Table G-84. P, R, F-Score for 30s

6. Extracted Test Data: 20s and 40s

20-29	Precision	Recall	F-Score
Baseline	0.751937984	1	0.85840708
1	0.748031496	0.979381443	0.848214286
2	0.818181818	0.092783505	0.166666667
3	0.751937984	1	0.85840708
4	0.751937984	1	0.85840708
5	0.751937984	1	0.85840708
6	0.751937984	1	0.85840708
7	0.751937984	1	0.85840708
8	0.751937984	1	0.85840708
9	0.751937984	1	0.85840708
10	0.857142857	0.06185567	0.115384615
11	0.75	0.989690722	0.853333333
12	0.7578125	1	0.862222222
13	0.751937984	1	0.85840708
14	0.75	0.989690722	0.853333333
15	0.833333333	0.051546392	0.097087379
16	1	0.041237113	0.079207921
17	0.75	0.989690722	0.853333333
18	0.75	0.989690722	0.853333333
19	0.751937984	1	0.85840708
20	0.751937984	1	0.85840708
21	0.751937984	1	0.85840708
22	0.755905512	0.989690722	0.857142857
23	0.751937984	1	0.85840708
24	0.75	0.989690722	0.853333333
25	0.751937984	1	0.85840708
26	0.763779528	1	0.866071429
27	0.748031496	0.979381443	0.848214286
28	0.748031496	0.979381443	0.848214286
29	0.751937984	1	0.85840708
30	0.751937984	1	0.85840708
31	0.75	0.989690722	0.853333333
32	0.75	0.989690722	0.853333333
33	0.75	0.989690722	0.853333333
34	0.75	0.989690722	0.853333333
35	0.751937984	1	0.85840708
36	0.75	0.989690722	0.853333333
37	0.751937984	1	0.85840708
38	0.751937984	1	0.85840708
39	0.755905512	0.989690722	0.857142857
40	0.7578125	1	0.862222222
41	0.751937984	1	0.85840708
42	0.751937984	1	0.85840708

43	0.751937984	1	0.85840708
44	0.751937984	1	0.85840708
45	0.751937984	1	0.85840708
46	0.751937984	1	0.85840708
47	0.751937984	1	0.85840708
48	0.751937984	1	0.85840708
49	0.751937984	1	0.85840708
50	0.7578125	1	0.862222222
51	0.751937984	1	0.85840708
52	0.751937984	1	0.85840708
53	0.75	0.989690722	0.853333333
54	0.751937984	1	0.85840708
55	0.751937984	1	0.85840708
56	0.748031496	0.979381443	0.848214286
57	0.751937984	1	0.85840708
58	0.751937984	1	0.85840708
59	0.7578125	1	0.862222222
60	0.7578125	1	0.862222222
61	0.746031746	0.969072165	0.843049327
62	0.818181818	0.092783505	0.166666667
63	0.751937984	1	0.85840708
64	1	0.06185567	0.116504854
65	0.751937984	1	0.85840708
66	0.751937984	1	0.85840708
67	0.751937984	1	0.85840708
68	0.7578125	1	0.862222222
69	0.751937984	1	0.85840708
70	0.75	0.989690722	0.853333333
71	0.751937984	1	0.85840708
72	0.751937984	1	0.85840708
73	0.75	0.989690722	0.853333333
74	0.753968254	0.979381443	0.852017937
75	0.751937984	1	0.85840708
76	0.751937984	1	0.85840708
77	0.805555556	0.298969072	0.436090226
78	0.757575758	0.257731959	0.384615385
79	0.751937984	1	0.85840708
80	0.751937984	1	0.85840708
81	0.751937984	1	0.85840708
82	0.751937984	1	0.85840708
83	0.751937984	1	0.85840708
84	0.751937984	1	0.85840708

Table G-85. P, R, F-Score for 20s

40-49	Precision	Recall	F-Score
Baseline	0.248062016	1	0.397515528
2	0.254237288	0.9375	0.4
10	0.254098361	0.96875	0.402597403
12	1	0.03125	0.060606061
15	0.25203252	0.96875	0.4
16	0.256	1	0.407643312
22	0.5	0.03125	0.058823529
26	1	0.0625	0.117647059
39	0.5	0.03125	0.058823529
40	1	0.03125	0.060606061
50	1	0.03125	0.060606061
59	1	0.03125	0.060606061
60	1	0.03125	0.060606061
62	0.254237288	0.9375	0.4
64	0.260162602	1	0.412903226
68	1	0.03125	0.060606061
74	0.333333333	0.03125	0.057142857
77	0.268817204	0.78125	0.4
78	0.25	0.75	0.375

Table G-86.

P, R, F-Score for 40s

7. Extracted Test Data: 20s and 50s

20-29	Precision	Recall	F-Score
Baseline	0.906542056	1	0.950980392
1	0.904761905	0.979381443	0.940594059
2	0.897959184	0.907216495	0.902564103
3	0.906542056	1	0.950980392
4	0.906542056	1	0.950980392
5	0.906542056	1	0.950980392
6	0.906542056	1	0.950980392
7	0.906542056	1	0.950980392
8	0.906542056	1	0.950980392
9	0.906542056	1	0.950980392
10	0.906542056	1	0.950980392
11	0.912621359	0.969072165	0.94
12	0.920792079	0.958762887	0.939393939
13	0.906542056	1	0.950980392
14	0.905660377	0.989690722	0.945812808
15	0.903846154	0.969072165	0.935323383
16	0.903846154	0.969072165	0.935323383
17	0.905660377	0.989690722	0.945812808
18	0.905660377	0.989690722	0.945812808
19	0.906542056	1	0.950980392
20	0.906542056	1	0.950980392
21	0.906542056	1	0.950980392
22	0.905660377	0.989690722	0.945812808
23	0.906542056	1	0.950980392
24	0.905660377	0.989690722	0.945812808
25	0.906542056	1	0.950980392
26	0.906542056	1	0.950980392
27	0.904761905	0.979381443	0.940594059
28	0.904761905	0.979381443	0.940594059
29	0.906542056	1	0.950980392
30	0.906542056	1	0.950980392
31	0.905660377	0.989690722	0.945812808
32	0.905660377	0.989690722	0.945812808
33	0.905660377	0.989690722	0.945812808
34	0.905660377	0.989690722	0.945812808
35	0.906542056	1	0.950980392
36	0.905660377	0.989690722	0.945812808
37	0.906542056	1	0.950980392
38	0.906542056	1	0.950980392
39	0.905660377	0.989690722	0.945812808
40	0.905660377	0.989690722	0.945812808
41	0.906542056	1	0.950980392
42	0.906542056	1	0.950980392

43	0.906542056	1	0.950980392
44	0.906542056	1	0.950980392
45	0.906542056	1	0.950980392
46	0.906542056	1	0.950980392
47	0.906542056	1	0.950980392
48	0.906542056	1	0.950980392
49	0.906542056	1	0.950980392
50	0.91509434	1	0.955665025
51	0.906542056	1	0.950980392
52	0.906542056	1	0.950980392
53	0.905660377	0.989690722	0.945812808
54	0.906542056	1	0.950980392
55	0.905660377	0.989690722	0.945812808
56	0.904761905	0.979381443	0.940594059
57	0.906542056	1	0.950980392
58	0.906542056	1	0.950980392
59	0.906542056	1	0.950980392
60	0.906542056	1	0.950980392
61	0.904761905	0.979381443	0.940594059
62	1	0.06185567	0.116504854
63	0.906542056	1	0.950980392
64	1	0.06185567	0.116504854
65	0.906542056	1	0.950980392
66	0.906542056	1	0.950980392
67	0.906542056	1	0.950980392
68	0.906542056	1	0.950980392
69	0.906542056	1	0.950980392
70	0.905660377	0.989690722	0.945812808
71	0.906542056	1	0.950980392
72	0.906542056	1	0.950980392
73	0.905660377	0.989690722	0.945812808
74	0.906542056	1	0.950980392
75	0.906542056	1	0.950980392
76	0.906542056	1	0.950980392
77	0.903846154	0.969072165	0.935323383
78	0.9125	0.75257732	0.824858757
79	0.906542056	1	0.950980392
80	0.906542056	1	0.950980392
81	0.906542056	1	0.950980392
82	0.906542056	1	0.950980392
83	0.906542056	1	0.950980392
84	0.906542056	1	0.950980392

Table G-87. P, R, F-Score for 20s

50-59	Precision	Recall	F-Score
Baseline	0.093457944	1	0.170940171
11	0.25	0.1	0.142857143
12	0.333333333	0.2	0.25
50	1	0.1	0.181818182
62	0.099009901	1	0.18018018
64	0.099009901	1	0.18018018
78	0.111111111	0.3	0.162162162

Table G-88. P, R, F-Score for 50s

8. Extracted Test Data: 30s and 40s

30-39	Precision	Recall	F-Score
Baseline	0.536231884	1	0.698113208
1	0.529411765	0.972972973	0.685714286
2	0.552238806	1	0.711538462
3	0.5	0.027027027	0.051282051
4	0.537313433	0.972972973	0.692307692
5	0.536231884	1	0.698113208
6	0.536231884	1	0.698113208
7	0.536231884	1	0.698113208
8	0.536231884	1	0.698113208
9	0.6	0.081081081	0.142857143
10	0.666666667	0.054054054	0.1
11	1	0.027027027	0.052631579
12	0.5	0.027027027	0.051282051
13	0.529411765	0.972972973	0.685714286
14	0.529411765	0.972972973	0.685714286
15	1	0.054054054	0.102564103
16	1	0.027027027	0.052631579
17	0.536231884	1	0.698113208
18	0.536231884	1	0.698113208
19	0.529411765	0.972972973	0.685714286
20	0.529411765	0.972972973	0.685714286
21	0.536231884	1	0.698113208
22	0.544117647	1	0.704761905
23	0.536231884	1	0.698113208
24	0.536231884	1	0.698113208
25	0.536231884	1	0.698113208
26	0.536231884	1	0.698113208
27	0.536231884	1	0.698113208
28	0.536231884	1	0.698113208
29	0.536231884	1	0.698113208
30	0.536231884	1	0.698113208
31	0.536231884	1	0.698113208
32	0.537313433	0.972972973	0.692307692
33	0.536231884	1	0.698113208
35	1	0.027027027	0.052631579
36	0.536231884	1	0.698113208
37	0.536231884	1	0.698113208
38	0.536231884	1	0.698113208
42	0.536231884	1	0.698113208
43	0.536231884	1	0.698113208
44	0.536231884	1	0.698113208
49	1	0.027027027	0.052631579
50	0.536231884	1	0.698113208

51	0.536231884	1	0.698113208
52	0.536231884	1	0.698113208
53	0.529411765	0.972972973	0.685714286
54	0.536231884	1	0.698113208
56	0.529411765	0.972972973	0.685714286
57	0.536231884	1	0.698113208
58	0.536231884	1	0.698113208
59	0.544117647	1	0.704761905
60	0.536231884	1	0.698113208
61	0.532258065	0.891891892	0.666666667
62	1	0.054054054	0.102564103
65	0.536231884	1	0.698113208
66	0.536231884	1	0.698113208
67	0.536231884	1	0.698113208
68	0.544117647	1	0.704761905
69	0.536231884	1	0.698113208
70	0.536231884	1	0.698113208
71	0.529411765	0.972972973	0.685714286
72	0.536231884	1	0.698113208
74	0.536231884	1	0.698113208
75	0.536231884	1	0.698113208
76	0.536231884	1	0.698113208
77	0.642857143	0.243243243	0.352941176
78	0.710526316	0.72972973	0.72
80	0.536231884	1	0.698113208
82	0.536231884	1	0.698113208
83	0.536231884	1	0.698113208
84	0.536231884	1	0.698113208

Table G-89.

P, R, F-Score for 30s

40-49	Precision	Recall	F-Score
Baseline	0.463768116	1	0.633663366
2	1	0.0625	0.117647059
3	0.462686567	0.96875	0.626262626
4	0.5	0.03125	0.058823529
9	0.46875	0.9375	0.625
10	0.46969697	0.96875	0.632653061
11	0.470588235	1	0.64
12	0.462686567	0.96875	0.626262626
15	0.47761194	1	0.646464646
16	0.470588235	1	0.64
22	1	0.03125	0.060606061
32	0.5	0.03125	0.058823529
34	0.463768116	1	0.633663366
35	0.470588235	1	0.64
39	0.455882353	0.96875	0.62
40	0.455882353	0.96875	0.62
41	0.463768116	1	0.633663366
45	0.455882353	0.96875	0.62
46	0.463768116	1	0.633663366
47	0.463768116	1	0.633663366
48	0.463768116	1	0.633663366
49	0.470588235	1	0.64
55	0.463768116	1	0.633663366
59	1	0.03125	0.060606061
61	0.428571429	0.09375	0.153846154
62	0.47761194	1	0.646464646
63	0.463768116	1	0.633663366
64	0.463768116	1	0.633663366
68	1	0.03125	0.060606061
73	0.463768116	1	0.633663366
77	0.490909091	0.84375	0.620689655
78	0.677419355	0.65625	0.666666667
79	0.463768116	1	0.633663366
81	0.463768116	1	0.633663366

Table G-90. P, R, F-Score for 40s

9. Extracted Test Data: 30s and 50s

30-39	Precision	Recall	F-Score
Baseline	0.787234043	1	0.880952381
1	0.782608696	0.972972973	0.86746988
2	0.804347826	1	0.891566265
3	0.787234043	1	0.880952381
4	0.782608696	0.972972973	0.86746988
5	0.787234043	1	0.880952381
6	0.787234043	1	0.880952381
7	0.787234043	1	0.880952381
8	0.787234043	1	0.880952381
9	0.772727273	0.918918919	0.839506173
10	0.782608696	0.972972973	0.86746988
11	0.8	0.972972973	0.87804878
12	0.822222222	1	0.902439024
13	0.782608696	0.972972973	0.86746988
14	0.777777778	0.945945946	0.853658537
15	0.787234043	1	0.880952381
16	0.787234043	1	0.880952381
17	0.787234043	1	0.880952381
18	0.787234043	1	0.880952381
19	0.782608696	0.972972973	0.86746988
20	0.782608696	0.972972973	0.86746988
21	0.787234043	1	0.880952381
22	0.787234043	1	0.880952381
23	0.787234043	1	0.880952381
24	0.787234043	1	0.880952381
25	0.787234043	1	0.880952381
26	0.787234043	1	0.880952381
27	0.787234043	1	0.880952381
28	0.787234043	1	0.880952381
29	0.787234043	1	0.880952381
30	0.787234043	1	0.880952381
31	0.787234043	1	0.880952381
32	0.782608696	0.972972973	0.86746988
33	0.787234043	1	0.880952381
34	0.787234043	1	0.880952381
35	0.782608696	0.972972973	0.86746988
36	0.787234043	1	0.880952381
37	0.782608696	0.972972973	0.86746988
38	0.782608696	0.972972973	0.86746988
39	0.787234043	1	0.880952381
40	0.787234043	1	0.880952381
41	0.787234043	1	0.880952381
42	0.787234043	1	0.880952381

43	0.787234043	1	0.880952381
44	0.787234043	1	0.880952381
45	0.787234043	1	0.880952381
46	0.787234043	1	0.880952381
47	0.795454545	0.945945946	0.864197531
48	0.782608696	0.972972973	0.86746988
49	0.782608696	0.972972973	0.86746988
50	0.782608696	0.972972973	0.86746988
51	0.787234043	1	0.880952381
52	0.787234043	1	0.880952381
53	0.787234043	1	0.880952381
54	0.772727273	0.918918919	0.839506173
55	0.787234043	1	0.880952381
56	0.782608696	0.972972973	0.86746988
57	0.787234043	1	0.880952381
58	0.787234043	1	0.880952381
59	0.787234043	1	0.880952381
60	0.787234043	1	0.880952381
61	0.777777778	0.945945946	0.853658537
62	0.76744186	0.891891892	0.825
63	0.787234043	1	0.880952381
65	0.787234043	1	0.880952381
66	0.787234043	1	0.880952381
67	0.787234043	1	0.880952381
68	0.787234043	1	0.880952381
69	0.787234043	1	0.880952381
70	0.787234043	1	0.880952381
71	0.782608696	0.972972973	0.86746988
72	0.787234043	1	0.880952381
73	0.787234043	1	0.880952381
74	0.787234043	1	0.880952381
75	0.787234043	1	0.880952381
76	0.787234043	1	0.880952381
77	0.852941176	0.783783784	0.816901408
78	0.818181818	0.72972973	0.771428571
79	0.787234043	1	0.880952381
80	0.787234043	1	0.880952381
81	0.787234043	1	0.880952381
82	0.787234043	1	0.880952381
83	0.782608696	0.972972973	0.86746988
84	0.787234043	1	0.880952381

Table G-91.

P, R, F-Score for 30s

50-59	Precision	Recall	F-Score
Baseline	0.212765957	1	0.350877193
2	1	0.1	0.181818182
11	0.5	0.1	0.166666667
12	1	0.2	0.333333333
47	0.333333333	0.1	0.153846154
64	0.212765957	1	0.350877193
77	0.384615385	0.5	0.434782609
78	0.285714286	0.4	0.333333333

Table G-92.

P, R, F-Score for 50s

10. Extracted Test Data: 40s and 50s

40-49	Precision	Recall	F-Score
Baseline	0.761904762	1	0.864864865
1	0.761904762	1	0.864864865
2	0.761904762	1	0.864864865
3	1	0.03125	0.060606061
4	0.761904762	1	0.864864865
5	0.761904762	1	0.864864865
6	0.756097561	0.96875	0.849315068
7	0.761904762	1	0.864864865
8	0.761904762	1	0.864864865
9	0.75	0.9375	0.833333333
10	0.761904762	1	0.864864865
11	0.761904762	1	0.864864865
12	0.8	1	0.888888889
13	0.761904762	1	0.864864865
14	0.780487805	1	0.876712329
15	0.761904762	1	0.864864865
16	0.761904762	1	0.864864865
17	0.761904762	1	0.864864865
18	0.780487805	1	0.876712329
19	0.761904762	1	0.864864865
20	0.761904762	1	0.864864865
21	0.756097561	0.96875	0.849315068
22	0.756097561	0.96875	0.849315068
23	0.761904762	1	0.864864865
24	0.761904762	1	0.864864865
25	0.72972973	0.84375	0.782608696
26	0.72972973	0.84375	0.782608696
27	0.775	0.96875	0.861111111
28	0.775	0.96875	0.861111111
29	0.761904762	1	0.864864865
30	0.761904762	1	0.864864865
31	0.761904762	1	0.864864865
32	0.761904762	1	0.864864865
33	0.761904762	1	0.864864865
34	0.761904762	1	0.864864865
35	0.761904762	1	0.864864865
36	0.761904762	1	0.864864865
37	0.761904762	1	0.864864865
38	0.761904762	1	0.864864865
39	0.756097561	0.96875	0.849315068
40	0.761904762	1	0.864864865
41	0.761904762	1	0.864864865
42	0.761904762	1	0.864864865

43	0.761904762	1	0.864864865
44	0.761904762	1	0.864864865
45	0.756097561	0.96875	0.849315068
46	0.756097561	0.96875	0.849315068
47	0.761904762	1	0.864864865
48	0.761904762	1	0.864864865
49	0.761904762	1	0.864864865
50	0.761904762	1	0.864864865
51	0.780487805	1	0.876712329
52	0.780487805	1	0.876712329
53	0.761904762	1	0.864864865
54	0.761904762	1	0.864864865
55	0.761904762	1	0.864864865
56	0.761904762	1	0.864864865
57	0.761904762	1	0.864864865
58	0.761904762	1	0.864864865
59	0.761904762	1	0.864864865
60	0.756097561	0.96875	0.849315068
61	0.761904762	1	0.864864865
62	0.75	0.9375	0.833333333
63	0.761904762	1	0.864864865
65	0.761904762	1	0.864864865
66	0.761904762	1	0.864864865
67	0.761904762	1	0.864864865
68	0.761904762	1	0.864864865
69	0.761904762	1	0.864864865
70	0.761904762	1	0.864864865
71	0.761904762	1	0.864864865
72	0.761904762	1	0.864864865
73	0.761904762	1	0.864864865
74	0.761904762	1	0.864864865
75	0.761904762	1	0.864864865
76	0.761904762	1	0.864864865
77	0.763157895	0.90625	0.828571429
78	0.692307692	0.5625	0.620689655
79	0.761904762	1	0.864864865
80	0.761904762	1	0.864864865
81	0.761904762	1	0.864864865
82	0.761904762	1	0.864864865
83	0.761904762	1	0.864864865
84	0.756097561	0.96875	0.849315068

Table G-93. P, R, F-Score for 40s

50-59	Precision	Recall	F-Score
Baseline	0.238095238	1	0.384615385
3	0.243902439	1	0.392156863
12	1	0.2	0.333333333
14	1	0.1	0.181818182
18	1	0.1	0.181818182
27	0.5	0.1	0.166666667
28	0.5	0.1	0.166666667
51	1	0.1	0.181818182
52	1	0.1	0.181818182
64	0.238095238	1	0.384615385
77	0.25	0.1	0.142857143
78	0.125	0.2	0.153846154

Table G-94.

P, R, F-Score for 50s

11. Extracted Test Data: Under 26 and 26 or Over

< 26	Precision	Recall	F-Score
Baseline	0.540983607	1	0.70212766
1	0.540983607	1	0.70212766
2	0.722222222	0.098484848	0.173333333
3	0.543209877	1	0.704
4	0.543209877	1	0.704
5	0.540983607	1	0.70212766
6	0.540983607	1	0.70212766
7	0.540983607	1	0.70212766
9	1	0.007575758	0.015037594
10	0.666666667	0.015151515	0.02962963
11	0.539748954	0.977272727	0.69541779
12	0.54893617	0.977272727	0.702997275
14	0.333333333	0.007575758	0.014814815
15	0.769230769	0.075757576	0.137931034
16	0.857142857	0.090909091	0.164383562
17	0.540983607	1	0.70212766
18	1	0.03030303	0.058823529
19	0.540983607	1	0.70212766
21	0.540983607	1	0.70212766
22	0.53909465	0.992424242	0.698666667
23	1	0.007575758	0.015037594
24	0.537190083	0.984848485	0.695187166
25	0.53909465	0.992424242	0.698666667
26	1	0.007575758	0.015037594
27	1	0.007575758	0.015037594
28	0.537190083	0.984848485	0.695187166
29	0.53909465	0.992424242	0.698666667
30	0.540983607	1	0.70212766
33	0.53909465	0.992424242	0.698666667
34	0.537190083	0.984848485	0.695187166
35	0.5	0.007575758	0.014925373
36	0.540983607	1	0.70212766
37	0.5	0.007575758	0.014925373
38	0.541322314	0.992424242	0.700534759
39	0.541322314	0.992424242	0.700534759
40	0.540983607	1	0.70212766
41	0.540983607	1	0.70212766
42	1	0.007575758	0.015037594
43	0.540983607	1	0.70212766
44	0.540983607	1	0.70212766
45	0.540983607	1	0.70212766
47	1	0.007575758	0.015037594
48	0.53909465	0.992424242	0.698666667

49	0.540983607	1	0.70212766
50	0.543209877	1	0.704
51	0.540983607	1	0.70212766
52	0.540983607	1	0.70212766
53	0.543209877	1	0.704
54	0.540983607	1	0.70212766
55	0.53909465	0.992424242	0.698666667
56	0.531380753	0.962121212	0.684636119
58	0.540983607	1	0.70212766
59	0.540983607	1	0.70212766
60	0.543209877	1	0.704
61	0.540983607	1	0.70212766
62	0.545454545	0.045454545	0.083916084
63	0.53909465	0.992424242	0.698666667
64	0.916666667	0.083333333	0.152777778
65	0.540983607	1	0.70212766
66	0.540983607	1	0.70212766
67	0.540983607	1	0.70212766
69	0.53909465	0.992424242	0.698666667
70	0.53909465	0.992424242	0.698666667
72	0.540983607	1	0.70212766
73	0.543209877	1	0.704
75	0.540983607	1	0.70212766
76	0.540983607	1	0.70212766
77	0.590909091	0.295454545	0.393939394
78	0.6	0.25	0.352941176
80	0.540983607	1	0.70212766
82	0.540983607	1	0.70212766
83	0.540983607	1	0.70212766
84	0.540983607	1	0.70212766

Table G-95. P, R, F-Score for Under 26

>= 26	Precision	Recall	F-Score
Baseline	0.459016393	1	0.629213483
2	0.473451327	0.955357143	0.633136095
3	1	0.008928571	0.017699115
4	1	0.008928571	0.017699115
8	0.459016393	1	0.629213483
9	0.46090535	1	0.630985915
10	0.460580913	0.991071429	0.628895184
11	0.4	0.017857143	0.034188034
12	0.666666667	0.053571429	0.099173554
13	0.459016393	1	0.629213483
14	0.456431535	0.982142857	0.623229462
15	0.471861472	0.973214286	0.635568513
16	0.47826087	0.982142857	0.643274854
18	0.466666667	1	0.636363636
20	0.456790123	0.991071429	0.625352113
23	0.46090535	1	0.630985915
26	0.46090535	1	0.630985915
27	0.46090535	1	0.630985915
31	0.459016393	1	0.629213483
32	0.459016393	1	0.629213483
35	0.458677686	0.991071429	0.627118644
37	0.458677686	0.991071429	0.627118644
38	0.5	0.008928571	0.01754386
39	0.5	0.008928571	0.01754386
42	0.46090535	1	0.630985915
46	0.459016393	1	0.629213483
47	0.46090535	1	0.630985915
50	1	0.008928571	0.017699115
53	1	0.008928571	0.017699115
57	0.459016393	1	0.629213483
60	1	0.008928571	0.017699115
62	0.459227468	0.955357143	0.620289855
64	0.478448276	0.991071429	0.645348837
68	0.459016393	1	0.629213483
71	0.456790123	0.991071429	0.625352113
73	1	0.008928571	0.017699115
74	0.459016393	1	0.629213483
77	0.47752809	0.758928571	0.586206897
78	0.476190476	0.803571429	0.598006645
79	0.459016393	1	0.629213483
81	0.459016393	1	0.629213483

Table G-96. P, R, F-Score for 26 or Older

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX H: KEY FOR FEATURE VECTORS AND FEATURES

This appendix contains the keys used in Appendix F for the feature vectors and Appendix G for the individual features.

A. KEY FOR FEATURE VECTORS

Tokens		Keys from Table H-2	Types		Keys from Table H-2
1	Emoticon	Odd keys: 19-34, 37-38, 45-46	2	Emoticon	Even keys: 19-34, 37-38, 45-46
3	Punctuation	Odd keys: 1-18, 35-36, 39-44. 47-76, 79-84	4	Punctuation	Even keys: 1-18, 35-36, 39-44. 47-76, 79-84
5	Word	77	6	Word	78

Table H-1. Key for Feature Vectors in Appendix F

B. KEY FOR FEATURES

Tokens		Types	
1	!	2	!
3	#	4	#
5	%	6	%
7	&	8	&
9	'	10	'
11	,	12	,
13	-	14	-
15	/	16	/
17	:	18	:
19	:-@	20	:-@
21	:- (22	:- (
23	:-)	24	:-)
25	:-o	26	:-o
27	:beer:	28	:beer:
29	:blush:	30	:blush:

31	:love:	32	:love:
33	:tongue:	34	:tongue:
35	;	36	;
37	i-)	38	i-)
39	<	40	<
41	=	42	=
43	>	44	>
45	>:->	46	>:->
47	@	48	@
49	"	50	"
51	\$	52	\$
53	(54	(
55)	56)
57	*	58	*
59	+	60	+
61	.	62	.
63	?	64	?
65	[66	[
67]	68]
69	^	70	^
71		72	
73	_	74	_
75	`	76	`
77	word	78	word
79	{	80	{
81	}	82	}
83	~	84	~

Table H-2. Key for Features in Appendix G

LIST OF REFERENCES

1. Argamon, S., Dawhle, S., Koppel, M., Pennebaker, J. Lexical Predictors of Personality Type. In Proceedings of Classification Society of North America, St. Louis MI, June 2005.
2. Argamon, S., Saric, M., Stein, S. Style Mining of Electronic Messages for Multiple Authorship Discrimination: First Results. In Proceedings of ACM Conference on Knowledge Discovery and Data Mining, 2003.
3. Baayen, R.H., Van Halteren, H., Tweedie, F.J. Outside the Cave of Shadows: Using Syntactic Annotation to Enhance Authorship Attribution. *Literary and Linguistic Computing*, 11(3):121-131, 1996.
4. Brizendine, L. *The Female Brain*. Morgan Roads Book, August 2006.
5. Corney, M. Analysing E-mail Text authorship for Forensic Purposes. Master of Information Technology Thesis, Queensland University of Technology, 2003.
6. Corney, M., De Vel, O., Anderson, A., Mohay, G. Gender-Preferential Text Mining of E-mail Discourse. In 18th Annual Proceedings of Computer Security Applications Conference, pp. 282-289, 2002.
7. De Vel, O. Mining E-mail Authorship. In Workshop on Text Mining, ACM International Conference on Knowledge Discover and Data Mining, Boston, MA, USA 2000.
8. De Vel, O., Anderson, A., Corney, M., Mohay, G. Multi-topic E-mail Authorship Attribution Forensics. In Proceedings Workshop on Data Mining for Security Applications, 8th ACM Conference on Computer Security (CCS), 2001.
9. De Vel, O., Corney, M., Anderson, A., Mohay, G. Language and Gender Author Cohort Analysis of E-mail for Computer Forensics. In Proceedings of Digital Forensic Research Workshop, Syracuse, NY, August 2002.

10. De Vel, O., Corney, M., Anderson, A., Mohay, G. Language and Gender Author Cohort Analysis of E-mail for Computer Forensics. In Proceedings of Digital Forensic Research Workshop, Syracuse, NY, August 2002.
11. Diederich, J. Kindermann, J., Leopold, E., Paass, G. Authorship Attribution with Support Vector Machines. Applied Intelligence, 2000.
12. Gary, A., Sallis, P., MacDonell, S. Software forensics: Extending authorship analysis techniques to computer programs. In Proceedings 3rd Biannual Conf. Int. Assoc. of Forensic Linguists (IAFL'97), pp. 1-8, 1997.
13. Herring, S.C. Gender and Democracy in Computer-Mediated Communication. Electronic Journal of Communication, 3(2), 1993.
14. Hills, M. You are What You Type: Language and Gender Deception on the Internet. Bachelor of Arts with Honors Thesis, University of Otago, 2000.
15. Hsu, C.W. and Lin, C.J. A Comparison of Methods for Multi-Class Support Vector Machines. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, 2001.
16. Internet Crimes Against Children, http://www.ojp.usdoj.gov/ovc/publications/bulletins/internet_2_2001/internet_2_01_6.html, December 2001. Last accessed March 18, 2007.
17. Jurafsky, D., Martin, J.H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice-Hall Inc., 2000.
18. Krusul, I. and Spafford, E. Authorship analysis: Identifying the author of a program. Computers and Security, 16:248-259, 1997.
19. Malyutov, M.B. Authorship Attribution of Texts: A Review. Submitted to Proceedings of Information Transfer, 2005.
20. Mascol, C. Curves of Pauline and Pseudo-Pauline Style i. Unitarian Review, 30:452-460, 1888.

21. Mascol, C. Curves of Pauline and Pseudo-Pauline Style
ii. Unitarian Review, 30:539-546, 1888.
22. Mendenhall, T. The Characteristic Curves of Composition.
Science, 214:237249, 1887.
23. Mosteller, F and Wallace, D.L. Inference and Disputed
Authorship: The Federalist. Series in behavioral
science: Quantitative methods edition. Addison-Wesley,
Massachusetts, 1964.
24. Ojemann, G. Brain Organization for Language from the
Perspective of Electrical Stimulation Mapping.
Behavioral and Brain Sciences, 6:189-230, 1983.
25. Rayson, P., Leech, G., Hodges, M. Social differentiation
in the use of English Vocabulary: Some Analysis of the
Conversational Component of the British National Corpus.
International Journal of Corpus Linguistics, 2(1):133-
152, 1997.
26. Russell, S.J., Norvig, P. Artificial Intelligence: A
Modern Approach. Pearson Education, Inc., 2003.
27. Savicki, V., Lingenfelter, D., Kelley, M. Gender
Language Style and Group Composition in Internet
Discussion Groups. Journal of Computer Mediated
Communication, 2(3), 1996.
28. Schler, J., Koppel, M., Argamon, S., Pennebaker, J.
Effects of Age and Gender on Blogging. In Proceedings of
2006 AAAI Spring Symposium on Computational Approaches
for Analyzing Weblogs, 2006.
29. Singh, S. A Pilot Study on Gender Differences in
Conversational Speech on Lexical Richness Measures.
Literary and Linguistic Computing, 16(3):251-264, 2001.
30. Sussman, N.M., Tyson, D.H. Sex and Power: Gender
Differences in Computer-Mediated Interactions. Computers
in Human Behavior, 16:381-394, 2000.
31. Thomas, R., Murachver, T. Predicting Gender from
Electronic Discourse. British Journal of Social
Psychology, 40(2):193-208, 2001.

32. Tsuboi, Y., Matsumoto, Y. Authorship Identification for Heterogeneous Documents. Master's thesis, Nara Institute of Science and Technology, 2002.
33. Yule, G.U. On sentence length as a statistical characteristic of style in prose with application to two cases of disputed authorship. *Biometrika*, 30:363-390, 1938.
34. Yule, G.U. *Statistical Study of Literary Vocabulary*. Cambridge, U. Press, 1944.
35. Wolak, J., Mitchell, K., Finkelhor, D. Internet Sex Crimes Against Minors: The Response of Law Enforcement. National Center for Missing and Exploited Children, <http://www.unh.edu/ccrc/pdf/CV70.pdf>, November 2003. Last accessed March 18, 2007.
36. Zhang, H., Sheng, S. Learning Weighted Naïve Bayes with Accurate Ranking. *Data Mining (ICDM'04)*, pp.567-570, 2004.
37. Zipf, G.K. *Selected studies of the principle of relative frequency in language*. Harvard University Press, Cambridge MA, 1932.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California
3. Dr. Luqi
Naval Postgraduate School
Monterey, California
4. Dr. Gang Qu
University of Maryland, College Park
College Park, Maryland
5. Ron Chen
Defense Manpower Data Center
Seaside, California
6. Ann Wharton
Department of Defense
Linthicum, Maryland
7. Dr. Kevin Squire
Naval Postgraduate School
Monterey, California
8. Dr. Craig Martell
Naval Postgraduate School
Monterey, California
9. Jane Lin
Department of Defense
Monterey, California